Privacy & Security Tiger Team
Public Hearing
<mark>Draft Transcript</mark>
December 9, 2010

Presentation

Judy Sparrow – Office of the National Coordinator – Executive Director
Good afternoon, everybody, and welcome to the HIT Policy Committee's Privacy & Security Tiger Team. This is a Federal Advisory Committee, which means there will be opportunity at the end of the meeting for the public to make comment, and just a reminder for members of the workgroup to please identify yourselves for our listening audience.

Let's go around the table and introduce members of the workgroup starting on my left.

Joy Pritts – ONC – Chief Privacy Officer
Joy Pritts, Chief Privacy Officer, ONC.

Peter DeVault – Epic Systems – Project Manager
Peter DeVault, Epic.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Wes Rishel, Gartner.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
John Houston, UPMC.

Deven McGraw – Center for Democracy & Technology – Director
Deven McGraw with the Center for Democracy & Technology.

Paul Egerman – Software Entrepreneur
Paul Egerman, Software Entrepreneur.

Gayle Harrell – Florida – State Representative
Gayle Harrell, State Legislator, Florida.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
David McCallie, Cerner.

Adam Greene – Office of Civil Rights – Senior HIT & Privacy Specialist
Adam Greene, Office for Civil Rights.

Judy Sparrow – Office of the National Coordinator – Executive Director
On the phone, Carol Diamond?

Carol Diamond – Markle Foundation – Managing Director Healthcare Program
Yes. Carol Diamond with Markle.

Judy Sparrow – Office of the National Coordinator – Executive Director
Any other workgroup members on the telephone? Okay. With that, I'll turn it over to Deven McGraw.

Deven McGraw – Center for Democracy & Technology – Director
I'm going to turn it over to Paul Egerman.

Paul Egerman – Software Entrepreneur

Good morning. I want to welcome you all to our hearing on patient matching. I want to start by thanking our good friends from ONC who did all the work to organize this, so thank you Judy Sparrow, and thank you, Joy Pritts, for all of your work. I also want to thank, of course, the members of the tiger team who are here and members of the tiger team who are on the telephone. I especially want to thank the members of the public who, on this bright, chilly day are here to listen to our hearing and also members of the public who might be listening over the telephone or on the Internet. We very much appreciate your interest in our work, and we want to tell you that there will be an opportunity for public comment at the end of the hearing. I hope that you will take advantage of that. We are very much interested in public comments, and we are very much influenced by those comments.

I want to take a quick second to explain the hearing and what we're trying to do here. The hearing is on this topic called patient linking or patient matching. This group, we call ourselves the tiger team, or another word for it is that we are a taskforce. We report recommendations to the HIT Policy Committee, and that is formed under ARRA (American Recovery and Reinvestment Act) in the section called the HITECH Provisions. Within our industry, this area is now affectionately called the meaningful use stuff, and so we are part of the meaningful use stuff. That is what we are all about.

Our interest in patient linking and patient matching relates to privacy and security, but it also relates to interoperability, which is one of the main topics, main sections of interest within the whole meaningful use stuff is interoperability. That's really what we are interested in. We're interested in making sure that we understand when information or data goes from one healthcare organization to another healthcare organization, that the patient is properly identified and correctly matched or linked to the rest of the data on either side. We're also interested in things that may relate to that, so we may be talking about issues related to what happens in effect on the source side of that data or on the other side that may impact that entire process.

In arranging this hearing, basically, we sent to the members—the people who will be testifying—a list of questions that they will be responding to, and we also sent them some information that was also sent to the members of the tiger team. But I'm not sure that this was actually sent out to all the members, to the public, and there is one area on this topic of universal patient identifier. I'm not sure whether or not it was sent out, so I want to make sure that I read to everybody the exact wording of what was sent out to the people who are testifying on this subject. It relates to a public law. Here's what we sent out.

It says, "1999 Public Law 105277 prohibited the U.S. Department of Health and Human Services (HHS) from using any of its appropriated funds to promulgate or adopt any final standard providing for or providing for the assignment of a unique health identifier of an individual until legislation is enacted specifically approving the standard."

You say what does that all mean? It basically means HHS is constrained from implementing a unique health identifier when this language is in effect, which basically says to change that requires an act of Congress, and so that's information that was presented to everybody. It doesn't prohibit us from speaking about it, but people need to know that that's the situation. What will be an outcome of our work will be recommendations. We may make recommendations over the long-term, but I think the members of the public should understand that we also are likely to make recommendations that are very sort of pragmatic and short-term recommendations. We're very much focused on meaningful use, specifically stage two and stage three of meaningful use where fundamentally we've got to write the regulations in 2011 and also in 2013. So a lot of our recommendations may appear to be incremental, pragmatic, and tactical related to those two areas.

After today's hearing, we will have a telephone conference of the members of the committee to review what we heard. That telephone call will be in the public, and it'll be at 10:00 eastern time tomorrow morning and so that people can participate in that, and you can sign onto the HIT ONC Website to find out the details about how to sign on to that. Then on Monday, December 13th, Deven and I will be giving a short summary of what we think we heard during today's session.

I want to talk a little bit about the logistics as to how we're going to run the hearing, but let me first pause and see, Deven, if you would like to add anything in terms of general comments.

Deven McGraw – Center for Democracy & Technology – Director
No, I think you did an excellent job, Paul. If you're going to go through logistics, go right ahead.

Paul Egerman – Software Entrepreneur
Great. Just to explain to everybody, Deven and I had a fierce battle over who would get to read the law, and so I won this time, but she will get—that was the promise—that next time she gets the opportunity to actually read the law, and I'm sure she could do a far more passionate job.

Deven McGraw – Center for Democracy & Technology – Director
…do that in my job, read the law.

Paul Egerman – Software Entrepreneur
Yes. Deven is an attorney. Here's how we're going to be doing the hearing. The hearing is set up in like four panels, though there are really three topics. The first panel is going to be on basically what is the problem, making sure that we understand as best as we can what are the patient matching, patient linking problems. The next two panels will be talking about what are the solutions, what are the ways that people try to solve this problem. Then the final panel will relate to trying to understand what approaches are there in other industries outside of healthcare to this issue, particularly in the financial industry where there might be things that we could learn and understand that would be helpful in terms of our understanding of this issue.

The people who are testifying, I'd like to again thank you again. They've produced excellent, really excellent written testimony, all of which is on the Website. People are encouraged to read it, and we thank you for doing that. Then we asked each one of them to give an oral presentation that is limited to five minutes. The most important thing I want to make sure I say to those of you who are presenting is that really is a hard stop at the end of five minutes. It's really critically important that the most important part of this process is actually the questions and the discussion, and we need to have enough time for that.

The way we're going to organize things is Deven and I will alternate. Deven will do the first panel, and she'll be friendly, and she'll introduce everybody and try to encourage the discussion. Then I'll be the bad guy, and my job will be to stop you at the end of five minutes. I just want to explain what would happen to you should you speak longer than five minutes. If you speak longer than five minutes in a hearing on patient linking, all of your comments will be attributed to somebody else. That is what'll happen. Then when you call to complain and ask us to fix it, we will fix it, but we will take forever to fix it, so that's what'll happen. You should not do that. In addition, you will be audited by the IRS, but other than that, that's all that will happen to you, so you have to stop at the end of five minutes, and that will be my job.

I think there'll be a timer on the screen. Is that correct, Judy? So you will see the timer. So I appreciate your understanding of that, and why don't we get started with the first panel, Deven?

Deven McGraw – Center for Democracy & Technology – Director
Terrific. I'm just going to do some very brief introductions of the folks on panel one because their very impressive bios are in the materials that are both available on the Website, as well as in the hearing room. We have Barbara Demster from HIMSS. We have Scott Whyte, who is the senior director of physician and ambulatory IT strategy with Catholic Healthcare West; Shaun Grannis, who is the director and principle investigator for dry ice at the Regenstrief Institute; Brad Malin from AMIA; and we have a testifier on the phone, Garland Land, with the National Association for Public Health Statistics and Information Systems. We'll go in that order.

Barbara, you can begin. Thank you.

Barbara Demster – HIMSS – Chair, Patient Identity Integrity Workgroup & Consultant
Thank you, and good morning. My name is Barbara Demster. It is three A's and no P in Demster. On behalf of HIMSS and the patient identity integrity workgroup, I would like to thank you for inviting us to participate in the hearings on this most important subject of patient linking.

This panel has been asked to define the problem and to address why accurate patient linking is needed. The ultimate goal is the accurate identification of patient and linking all related information to that individual within and across systems. Linking the wrong clinical information to a person not only can cause great personal harm to the patient, but also can incur unnecessary costs to the healthcare provider in correcting and mitigating the error.

Incorrect information impacts patient safety and compromises quality of care. Good clinical decisions based on bad data become bad clinical outcomes. The wrong patient who receives the wrong implant may undergo a second procedure to correct it. A third procedure is performed on the correct patient who never got it due to the identity error.

Providers are negatively impacted by unnecessary costs incurred by the mistakes, including legal costs or damage to reputation. Patient identity integrity is needed because it has major impact on quality of care, patient safety, and an organization's financial bottom line. It is of particular importance that it be addressed now due to the impending exacerbation of linking issues with the movement towards health information exchange.

In February of 2008, key industry sources asked HIMSS to shed light on challenges relating to patient identity integrity and the need for guidance for industry and understanding the complex issues surrounding it. In response, HIMSS established a patient identity integrity workgroup and invited subject matter experts representing a cross section of stakeholders to participate. Representatives from major provider groups, vendors, and government agencies participated. The workgroup delivered a white paper in December of 2009 that's included in your packet, and that enhances the industry's understanding of patient identity integrity as a holistic business process.

The white paper deals with the step-by-step process for matching records on … individual within and across multiple systems. The workgroup again composed of subject matter experts pool their collective knowledge and identified nine major variables that influence our ability to build and sustain a database in a high state of identity integrity. These nine variables are listed in the report as standards, system interfaces, algorithms, unique identifiers, business processes, data accuracy, data quality, training, and medical devices.

Each of these has an impact on record linking and no one of them provides a magic solution. The white paper makes a total of 56 recommendations across the nine influencers. The executive summary of the white paper describes the major issues for each of these influencers, and the specific recommendations can be found in body of the report. Sixteen of the 56 recommendations related to medical devices, and these have been forwarded to the HIMSS medical device taskforce for follow through.

The patient identity integrity workgroup is now working on operationalizing the remaining recommendations. This includes public policy and industry activity, as well as development of a public resource through the HIMSS privacy and security toolkit. The workgroup's work plan for 2011 is included in the document submitted to the team.

The workgroup believes there is no one answer or perfect solution. Record linking is a complex task riddled with interdependent activities. All nine influencers must be addressed to move to a higher state of identity integrity. When one area fails, it impacts downstream processes. Patient identity integrity carries great impact on critical processes in the practice and business of medicine. We must aim for a set of best solutions that will improve the rate of successful matches and results in linking of information. Again, I thank you for this opportunity to address this important topic.

Deven McGraw – Center for Democracy & Technology – Director

Barbara, you did an excellent job.

M
She was off by one second.

Deven McGraw – Center for Democracy & Technology – Director
Yes. Thank you very much. Scott?

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
On behalf of our 41 hospitals in Arizona, California, Nevada, Catholic Healthcare West appreciates the opportunity to submit testimony to the policy committee's tiger team. CHW is the eighth largest hospital system in the nation and the largest hospital provider in California with facilities in California Arizona, and Nevada. I know that the tiger team values the important perspective of the community hospital, and I thin CHW's experience is reflective of much of the provider community across the country. In addition to our 41 hospitals, CHW is proud of our network of over 10,000 affiliated physicians. The vast majority of our facilities are not integrated with owned physician groups. This means that in addition to investing in our hospital information systems, CHW has to provide connectivity with physician partners.

CHW's communities are all unique, ranging from rural to downtown urban settings. CHW is making tremendous investments in advanced health IT while absorbing enormous costs associated with the implementation of a number of other initiatives, including meaningful use of ICD-10. All of these initiatives amount to cost, workflow changes, policy and procedure updates, and the retraining of tens of thousands of employees. All the while, the hospitals are struggling to keep the doors open. We experience the economic costs of imperfect patient linking, and we're concerned about the growing cost related to this topic.

CHW believes that the patient linking is one of the most important areas of focus within health IT policy and that the technology alone does not exist to bring patient identification to the level of accuracy that is foolproof. Fundamentally, accurate patient matching is a human driven process, dependent on the education, training, and commitment of all persons in the chain from the frontline receptionists to executive leadership, including employers and insurance providers. Though some technologies do exist to identify potential records to reconcile or link, many decisions for merging or linkage ultimately must be made by a person. Since a technology silver bullet does not exist, CHW looks to the ONC to encourage and facilitate the development of a solution framework to patient matching that is aligned with health outcome goals of improving care quality, improving access to care, and reducing the cost of care.

Our experience generally mirrors the national duplication rate of 8% to 12%, though we do have some facilities with much lower error rates correlating to our most sophisticated systems. Patient matching becomes proportionately more difficult the larger the entity is and the more separated it is from the patient, so HIE presents an even bigger challenge, both in the case of directed exchange where there's push and the patient is known, and in the case of query models for HIE.

I've highlighted a case study in my written testimony. There we see the example of a health information organization, which CHW is a participant. This HIO has been in existence for 14 years and, in the summer of 2009, started an investigation for how to clean up the demographic information. The team consisted of a project lead, an analyst, subject matter expert, just a full-time cleanup person, and then additional site administrators, engineers, and vendor representatives. If we apply that level of effort to the whole state of California, it is two orders of magnitude larger, so we're talking 100-fold increase in effort. You could argue that the problem in California is actually bigger than a 100 times larger because the patient matching issue increases in complexity, as the size of population grows and the distance from the points of data entry grow. The resources needed to match patient identities without an improved framework or common identifier in the state the size of California are absolutely massive.

In the year since the project was started, we have seen dramatic improvements with the nearly 50% drop in the suspected duplicates from 135,000 to 74,000, and a 70% drop in people who share a social security number from 41,000 to 17,000. As the ONC develops recommendations, CHW urges it to

recognize the fundamental need that human intervention should be enhanced by technology solutions. I sincerely appreciate the opportunity to submit this testimony. Thank you.

Deven McGraw – Center for Democracy & Technology – Director
Thank you very much. I just want to interrupt for a second. We're joined by another member of the tiger team, Alice Brown from the National Partnership for Women and Families. Go ahead, Shaun.

Shaun Grannis – Regenstrief Institute – Director & Principal Investigator, ICEPHI
Thanks. Good morning. I want to thank the committee for the opportunity to engage in the topic of vital importance that we're all well aware of. My brief testimony will cover, at a very high level, what goes, what more detail is provided in the packets previously supplied to you, so I'm going to start by answering the question specifically posed to this panel, which is, what's the problem, and why is accuracy important? Clinical data are scattered across many independent databases and systems where data is stored as separate islands with different patient identifiers. The situation impedes aggregation of information about individuals for many healthcare use cases, routine delivery of care, population health and public health level processes, etc. Therefore, both accurate and robust linkage capabilities and effective methods for cleaning and transforming low quality clinical data are needed.

We're talking about patient linkage today, but the intimately related partner here is data quality. We can't talk about linkage without talking about data quality. Our experience and the experience of others strongly indicate that improved data quality dramatically enhances patient linkage accuracy. An apocryphal maximum suggests that when given the choice between good data or a good algorithm, choose good data every time. Unfortunately, data shared through health information exchange is often suboptimal. HIEs pose particularly difficult record linkage challenges because they combined information that have different degrees of data quality varying both by data source and time. It's all over the place.

It's also important to keep in mind context. Context matters. The specific use case for which the linkage is done impacts the linkage accuracy. I'm going to give you two examples. For example, a large HIE supporting patient care will receive hundreds of thousands of transactions daily that all need to be matched, and that HIE may lack resources to manually review these questionable links. Therefore, to minimize human review of potential false positive linkages, you tune that algorithm to a very low false positive rate.

On the other end of the spectrum, untreated active tuberculosis patients in a community pose a great public health risk. To identity these individuals against a known list of untreated patients, a linkage approach that minimizes the false negative rate may be desired for a point of care clinical reminder delivered in an emergency department setting. So there's also a paucity of well-described literature around this.

The technology is mature. The theory is mature, but there have been few formal, comprehensive, peer review evaluations of the performance capabilities of patient linkage in this very mess, low quality data. Evaluations haven't reflected the challenges that we face. Often in the literature, the data goes through extensive preprocessing, extensive cleanup, or gathered through one single workflow, which does not reflect the realities of the data.

I'm going to talk very briefly about the notion of a universal healthcare identifier. We can't not talk about the elephant in the room. An identifier can improve patient linkage accuracy, though it's not a panacea because the identifier will not be universally available. It won't be recorded with perfect fidelity. And it might be used erroneously or fraudulently. Therefore, robust linkage algorithms are still going to be necessary, even in the era of unique identifiers.

Given that, the degree to which a unique identifier will help is entirely unclear. I would claim no one in this room or this country knows how well a national unique identifier will help, and that question needs to be answered if we're going to contemplate billions of dollars in deploying such a system. So what should ONC do to address patient matching problems in information exchange? In the document I've provided, there are at least eight recommendations.

I'm going to talk about three here. One, facilitate a process for describing current approaches to patient linkage including the technology used, the human resources used in that process, and what workflows it's applied to. Number two, foster consistent approaches to evaluating and disseminating the accuracy of various strategies, so if one is sharing how we do it, two is describe how well we do it. Then, three, it's unclear whether these different systems with their different approaches to patient matching can actually share data with each other. We don't know if there's going to be an impedance mismatch between them. As health information exchange expands, we're going to need to understand how existing solutions can work together. Thank you for your time.

Deven McGraw – Center for Democracy & Technology – Director
Thanks very much, Shaun. Brad?

Brad Malin – American Medical Informatics Association (AMIA)
Good morning and thank you for the opportunity to present testimony this morning. On behalf of the American Medical Informatics Association or AMIA, I am pleased to provide these remarks regarding important issues about patient matching, specifically those most relevant to clinical care and research. I currently work as an assistant professor of biomedical informatics in the school of medicine, and an assistant research professor of computer science in the school of engineering at Vanderbilt University, so I'm coming at you from the computer science, as well as the informatics side.

In the short amount of time available this morning, I'll highlight several key challenges and implications regarding patient matching. I've organized my remarks to address several overarching topics. First, in working with large data sets, one of the implications of misidentified data, that is, of being uncertain whether a particular data subject is accurately identified or repeated. The impact of misidentification itself is actually dependent on the purpose for which the linkage is being performed. To illustrate the differences, I'll take a moment to speak about two environments: patient care and biomedical research.

In the context of clinical care, a false match can cause a patient's record to accumulate erroneous information and can alter clinical care. We already heard some of that earlier today. By contrast, a false non-match results in incomplete knowledge about a patient. This can lead to harm if clinical information, such as when medications that could cause adverse events are not reported. Those such iterations could also lead to unnecessary replication of diagnostics or services, thus increasing the cost of care. Regardless of the impact, it's important to recognize that dirty data in the clinical realm may lead to clinicians distrust in historical information electronic medical records that they find in the patient record themselves.

In the biomedical research environment, false matches can have affect the system in several ways and depend on the action that results from the match. For example, if the action taken as a result of the record matches to delete one of the records from the system, this can lead to bias in any particular study. This bias can be positive, such as when the record contains an association with the clinical concept under inquiry, or it can be negative, such as when the record implied there was no particular association. On the other hand, false matches could lead to improper associations in the system if the records are integrated themselves and not deleted.

Second, are there particular kinds of research studies hampered that are more or less by the data matching problem? I just want to make clear that any type of research study can be affected by problems in patient matching. But in particular, longitudinal studies are of significant concern. In particular, a fragmentation problem is front and center. For instance, imagine you've got a subject who gets diagnosed with a bacterial infection at one institution, but it pronounced as deceased at a second institution, but you didn't integrate their records. In this situation, there's a potential association between bacterial infection and an increased likelihood of mortality that you're missing.

Third, what percentage of unique individuals in a large data set are incorrectly identified or characterized. How does this percentage vary across different types of databases? So data quality, as we've already heard, has a direct impact on patient matching accuracy. But different types of data sources end up

marrying with respect to the percent of incorrect identifications that exist. The exact numbers depend on what the semantics are in which the information is being derived. So I don't have full experience to speak about all the different kinds of databases and data sets that exist, but let me give several different examples from the electronic medical record in an epidemiological setting.

First, I'll mention a study that was actually instituted by Shaun Grannis in which they tried to link newborn screening records with medical records in the Indiana Network for Patient Care. They used state of the art probabilistic matching algorithms, and they basically found that 90% of the records that should be claimed as a match were in fact claimed as a match, whereas they found that 95% of the records that should not have been claimed as a match were actually correctly left in that state. He could speak about this in greater detail.

I'll also mention a second study that was done at the Washington State Division of Alcohol and Substance Abuse. They used 600,000 records with approximately 26% duplication, and they found that off the shelf record linkage software was able to achieve approximately 96% to 97% patient matching accuracy.

Next, I'll speak on how does the matching problem differ from other data quality issues, and basically I'll say that the matching problem is fundamentally a data quality issue. Patient record matching algorithms and software designed for it are basically error correction techniques. The more error ambiguity that you have in the system, the more difficult it's going to be to derive the true signal.

That said, there are significant problems that reside in this area. First, I'll just note that it's not uncommon for family members to share information from when reporting in the hospital, so you see social security numbers duplicated significantly. You also see social security numbers reported as medical record numbers. So there are a lot of other issues that I could address, but I'll allow you to read the testimony. On behalf of AMIA, I want to thank ONC and the HIT Policy Committee for your attention on this important public policy issue.

Deven McGraw – Center for Democracy & Technology – Director
Thank you very much. Garland Land, are you on the phone?

Garland Land – NAPHSIS – Executive Director
Yes, I am.

Deven McGraw – Center for Democracy & Technology – Director
Terrific. You can begin your testimony. As you heard earlier, you have five minutes, and my enforcer, who is sitting next to me, will give you a hint when you're close to the end.

Garland Land – NAPHSIS – Executive Director
Thank you very much. What I want to do is inform the committee on the type of linkages that are occurring in public health agencies today. There are really four different types that I'll be describing. The first is the need to link birth and death records. This requirement is needed to prevent persons from getting a birth certificate on a deceased person in an attempt to assume the false identity. Birth records of deceased persons are flagged so either the birth records are not issued or they're issued with a notation on the record that the person is deceased. This is important to prevent identity fraud.

Most states … their birth and death records through a computer algorithm, usually with a deterministic matching process that matches on name, date of birth, sex, and possibly parent's names. For more recent births, a social security number may also be available. There are obvious issues in performing this linkage. The names on the birth certificate and death certificate may be different. There could be an error on the date of birth on a death certificate, and there could actually possibly be a gender change.

Some states use a computer algorithm that requires a one-to-one match on the each variable. Actually, most states do this. The death records that don't match a birth record, a second pass or third pass is done with less restrictive criteria. If there's a one-to-many match or no match, then states rely on manual matching processes. Manual matching is not only required on 20% to 40% of the records, because of the

sensitive nature of mismatching a person as dead when they are alive, it is essential that a very conservative process be used in computer and hand matching. Small states totally rely on manual matching.

The second type of linkage occurs when matching deaths with a public health registry such as cancer, TB, AIDS, etc. The purpose is to pure the registry of death … longevity studies, matching variables similar to those used with birth records. In some states, the social security number may be available on the registry file that it is on the death file. The computer matching algorithm deployed is also similar to the one previously described depending on the size of the files. As with birth matching, it's important to keep the mismatches to a very low level. Of course, with the meaningful use having inundation data coming in from clinical records into the immunization register with the state, this matching is going to be of even more importance in the future.

The third type of linkage is what can be referred to as program integration. Birth records are used to establish a population base for immunization registers, newborn screening programs, newborn hearing programs and other child help programs. The intent is to establish the population that should be receiving services based upon the births in the state. Some states have a fully integrated child health system that links all child health program services in a single information system.

There are a wide variety of mechanisms used to link such files. Some states link the newborn screening and hearing information with the birth record at the time that the birth record is created in the hospital. This is done by creating a number in the hospital that is entered into the newborn screening, newborn hearing, and birth certificate systems. Eleven states actually have this type of approach, but only four of them require that number to be placed on the birth certificate.

The number then can be used to link the systems. In other states, separate databases are created for each of these programs and then linked later on through some type of computer algorithm. Some states have created a central patient index that is used to update records with current information on immunizations, lab results, and follow-up information. Besides the name, date of birth, and gender, the hospital at birth is often used to link these records. As in the previous example, it's critical to not have mismatches, so hand matching is used when there is an ambiguous computer match.

The fourth and final matching example is for statistical research purposes. Some states link sibling birth records to a mother to understand the behavior of health changes for all pregnancies of a woman. Other linkages are done for program evaluation purposes. For example, many states have linked the birth and WIC records to evaluate the WIC program. Usually such linkage activities do not require 100% matching, and it's less important to prevent mismatches. Probabilistic matching algorithms are often used for this purpose.

In conclusion, matching a birth, death, service, and patient records are an important part of public health activity. In some cases, it is important to link 100% of records with no mismatches or very few mismatches. In other cases, such a high degree of precision is not essential. The matching is usually done both by computer algorithm and through visual matching. Thank you.

Deven McGraw – Center for Democracy & Technology – Director
Thank you very much. We appreciate it. We'll now move into the question phase, and I see John's hand. You guys can also use the tankards as well to let me know that you want to ask. Since I saw John right out of the corner of my eye, go right ahead.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
Thank you very much. I think the testimony and the written documents are really fantastic, and knowing that we're trying to understand the problem, and I know we're going to have discussions about the solutions later, I did want to ask one question about the problem. Clearly, there's a very grave concern regarding making sure we get the information correct for the purposes of things like patient safety, quality of care. One of the things that's, I think, still within the purview of this committee, and actually it's in the title of the committee, is privacy. I would be interested in, and I think we all assume that obviously the

more accurate the match, the better the patient privacy. But can any of you speak to specific privacy or concerns associated with privacy and whether you've seen any type of response from patients not wanting to participate due to concerns over privacy because of patient matching issues and errors?

Shaun Grannis – Regenstrief Institute – Director & Principal Investigator, ICEPHI
My very short answer is I know of no patient in our information exchange that said I don't want to participate because your patient matching system concerns me.

Barbara Demster – HIMSS – Chair, Patient Identity Integrity Workgroup & Consultant
The first thing that comes to mind is the increasing refusal of people to share their social security number. I doubt that patients are aware of the matching that's going on behind the systems, but there is a reluctance. That has become such a popular hot point for identity theft that we see more and more reluctance to share information that has, in the past, been used quite actively in the matching process.

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
I have not heard an instance where a patient has not opted in, if you will, or chosen not to share information because of a concern with patient matching. I don't think the public has a high awareness of the degree of the problem here. Our systems generally err on the side of preventing false positives, so we have typically far more duplicate instances of records, which prevent disclosure, inappropriate disclosure of private information.

M
One of the premises to your question said that increased accuracy leads to better patient privacy on the heels of that. We also fiercely defend against the false positive match. But in reality, poor matching systems actually may provide greater privacy than better ones, right? Just clarifying the premise there.

Paul Egerman – Software Entrepreneur
You're right.

M
Yeah.

Garland Land – NAPHSIS – Executive Director
From the public health perspective, I have not heard of a concern. Oftentimes these linkage projects are done through an institutional review board process. Actually, the opposite occurs. Sometimes public health programs will send out notices to families about services, and if we haven't done the linkage, and a child has died, then we certainly hear about that, so they're actually wanting us to purge records of children or other people who have died.

Brad Malin – American Medical Informatics Association (AMIA)
I don't believe that you'll find any studies that say people are hesitant to give up information or share information due to the patient matching problem, but there is a lot of evidence to suggest that people go to different places for care to avoid certain purposes or certain instances. So there are lots of examples where people who are employed at one particular healthcare organization will go to a completely different healthcare organization so that they don't end up having their information looked at by their friends or relatives. So I don't think that there really is enough study to suggest at this time that the problem is that people don't care about privacy and linkage across disparate systems.

M
I think it's acceptable to say that this really isn't prominently a privacy issue. I just want to make sure that we—we haven't really talked about it, and I want to see if there were privacy implications that we put them on the table as well.

Deven McGraw – Center for Democracy & Technology – Director
Wes? I'm not sure I agree, but Wes?

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
First off, I want to commend the panel on bringing up a point that I learned in the 1980's when I did a patient matching system and put it in a small hospital, a rural hospital, and a large city hospital. Part of this system, it recorded what probability of match the clerk accepted as a match, and identified by clerk which ones were just grabbing the top one and not paying attention to the probability. We found that even in an academic hospital that it was very, very hard to change that behavior, that the goal was to put the first two letters of the last name in, and the first letter of the first name, get a search, pick the top one, and go. So I think Shaun's proposal of creating a feedback loop around the quality of matching first, I think, on the basis of a study. I hope I'm not misquoting you, Shaun, but first in terms of a way to study it, but also in terms, I think, of a longer-term requirement for meaningful use or some other thing, I think, is critical beyond the technology to actually achieving a better quality of patient matching.

I wanted to disagree a little bit that a bias in the system towards reducing false positives would protect privacy. It's been my impression that the people who are wanting to breach privacy or people who just want to do matches of people for purposes that are not necessary evil, but not necessarily … to the patient like publishers and so forth. They're much more tolerant of false negatives than we are in healthcare. So, to the extent that they have their own sources of data, their own algorithms, they're not really relying on us for the linkages, so I think that there are many scenarios where that's not true. Obviously somebody who goes to one institution and pays cash so that their stuff stays out of a medical record, that's a case where your prescription, Shaun, might help, but on a kind of iffy basis.

Scott, you talked about a difference in accuracy in your hospitals, and I think you said it was based on technology and perhaps some training or something going on around it. Could you just elaborate on that a little bit? You said you had, I think, 8% to 12% in general, but specific hospitals are much better. Is that right? Yes.

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Since we have 41 hospitals, and they have different, many, many different legacy systems, some are 30 years old. They're from the 80s, Wes.

Paul Egerman – Software Entrepreneur
… legacy system ….

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Right, right. Well, yes, and some of those do not have a sophisticated searching for the registration clerks, for instance. However, some of those hospitals may have very, very attentive medical record staff. The accountability training that allows them to drive relatively low duplication rates. So it is both technology and process. As we implement new system, so as we replace those old systems with the most current generation of systems, we typically perform a medical records cleanup effort where we bring in the most advanced software, and we go through, identify issues with probabilistic matching techniques, and then perform the manual effort to typically merge or unmerge, and correct the records. So it's a combination of people, process, and technology.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
But better technology is a key factor in it?

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Better technology is absolutely a huge help. Yes.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
At the risk of beginning to look at solutions, I think we have experienced already with matching that happens at the time of query versus matching that builds up and stores linkages among records. So essentially a matching system that says this patient is identified by this number in this system and that number in that system, and maintains that linkage. Is there any thinking in terms of the relative benefits of the two approaches? I'm, in particular, concerned that correcting false matches is difficult when there's stored linkages in various places.

Brad Malin – American Medical Informatics Association (AMIA)
There are challenges. There is definitely thinking along these lines in terms of determining which one will work best. One of the problems though with doing record linkage on the fly is the computational capability of the system. Some of the algorithms that are currently on the market or are designed are designed to be run overnight or completely distributed across extremely powerful machines because, when you're talking about large systems, you're talking about potentially a million by a million records. If you don't have a good blocking algorithm or blocking component of your algorithm, you're going to be waiting for a very long time until you get the correct result. If you want a quick and dirty answer, then sure, you might be able to go that route. But a lot of people, if they want more accurate results, they're willing to wait a longer period of time to get the correct match.

Shaun Grannis – Regenstrief Institute – Director & Principal Investigator, ICEPHI
I guess when I think about the query versus the accumulation over time, the thought that comes to mind is there's no free lunch here, so you're either establishing identity when you run the query, or you're doing it as new transactions come in. I think the fundamental question is not necessarily the timing of establishing identity, but the architecture you use to maintain linkages between identities. The system we've built maintains individual identities from every data submitter. So if a patient is seen at 13 different institutions in our information exchange, they have 13 records in the system. If there's a problem in the algorithm that accumulates over time, we can actually relink every single person, every single identity in the system. So whether I'm querying in real time, or whether I'm accumulating that knowledge, that linkage over time, I don't see a fundamental distinction there, and I might be missing something, but to my mind it's not about when the query is done. It's about how facile is the system to, how flexible is the system to update and correct what might be incorrect linkages.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Thanks. I would just note that SureScripts maintains no linkages for their pharmacy database, and I think a lot of what we're hearing here is something that several of you alerted us to at the start of the call, which is that the use cases argue for different solutions. There is no universal approach to patient linking. Thanks.

Deven McGraw – Center for Democracy & Technology – Director
Thank you. Peter?

Peter DeVault – Epic Systems – Project Manager
Much of the discussion with linkage is often about how errors are created and errors are propagated, but I think it's also interesting to think about how do corrections get propagated and what components of systems and architectures of systems improve the ability for corrections to be propagated. For example, Scott, when you discover a duplicate patient record and that gets corrected, how can that get propagated to other systems, and are there ways of designing systems in health information exchange such that the propagation rates can be improved? The question is for anybody.

Shaun Grannis – Regenstrief Institute – Director & Principal Investigator, ICEPHI
I'll take a first crack at that. Technically propagating corrections is fairly straightforward. Practically speaking, our algorithm frequently identifies duplicate patients from our sending institutions. They want nothing to do with that information because it's not a part of their workflow today. I think, in the future, there will be tighter coordination between the system. The information exchange's job in Indiana is to simply insure that you're aggregating clinical data across sources as accurately as possible. Our job is not to get into individual institutions' workflows. We've tried. In fact, one of the things that we do in this notion of correcting or propagating corrections is laboratory data notoriously is designed to get back to a provider, not back to a patient.

There's an assumption that the provider will know who John Smith is with very limited identifiers. So in the context of notifiable condition reporting to public health, public health needs a positive lab result. They need to know who that patient is. The data coming in from that transaction is limited, so we actually have built a process to enhance the data with the information contained in the information exchange so

that public health knows who that individual is who has the positive chlamydia that's, by law, reportable. So we can propagate corrections or enhancements to some partners because they want it. Other partners don't.

Barbara Demster – HIMSS – Chair, Patient Identity Integrity Workgroup & Consultant
One of the findings of the patient identity integrity workgroup was the lack of understanding of the importance of the business processes in resolving true identity errors, and that the successful organizations did maintain an in-house function that actively managed the processes. It's a people rich process beyond the technology to get to the desired result. The concerns of why we started it was the lack of basic understanding and attention to the business processes for data accuracy, data quality, monitoring, that sort of thing. The HIMSS security survey of this year, one of the questions was, "Does your institution have an identity management process?" Fifty-two of the respondents from hospitals said yes, which means that basically half of the hospitals don't even have a management system in house.

Deven McGraw – Center for Democracy & Technology – Director
Was that Carol? Hold on a minute. Do you want to get in the question queue?

Carol Diamond – Markle Foundation – Managing Director Healthcare Program
I just want to get in the queue. Thanks.

Deven McGraw – Center for Democracy & Technology – Director
You're in. Go ahead.

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
I don't have an answer for this, but I do think it absolutely highlights the challenge and kind of this network affect, if you will. Within a hospital, of course, there are multiple stores for patient data, and just keeping those multiple stores clean is a challenge. It takes a lot of manual effort. When you have the hospital plus, let's say, a group of independent physicians, it's more difficult for those independent physicians. They're not directed by the same management team, etc. They have different policies. Oftentimes they don't have the same economies of scale to have staff that do nothing but pay attention to those types of issues. If we then scale that to the broader community and include homecare and SNIF and other providers, local county facilities, etc., the problem is that much more difficult. Then, of course, if you scale that on a statewide level, very, very, very difficult, so I think the problem of correction gets harder and harder the further away from the source of the data you are.

Deven McGraw – Center for Democracy & Technology – Director
Thanks.

W
David … a long time.

Deven McGraw – Center for Democracy & Technology – Director
Why did I not see that? I'm sorry. Yes, David McCallie.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
I'm worried that you might have a right hemisphere neglect syndrome.

Deven McGraw – Center for Democracy & Technology – Director
Are you disclosing my health data?

David McCallie – Cerner Corporation – Vice President of Medical Informatics
No. Just making a curb shot diagnosis. I know that this panel is focused on definition of the problem more than the solution, but a lot of you have expertise in solutions to the problem, so I'm going to ask that question. In particular, I wonder, are the algorithms necessary to do the best possible, probabilistic matching, which it seemed in the short-term will be a required part of our solutions given that we don't have a UPI? Are the algorithms that do the best job of probabilistic matching, well understood, well

documented, and available, or is that an area where additional perhaps public funded research or collaborative work on a reference implementation would be meritorious?

Barbara Demster – HIMSS – Chair, Patient Identity Integrity Workgroup & Consultant
That was one of our recommendations from the white paper that there be some level of independent study to establish the effectiveness of algorithms because we truly don't know. We don't see that algorithms will ever be eliminated because they provide such a valuable process. In terms of the additive value of the UPI, there are indications and data that show that they have added value to the algorithm process. So it may be a hybrid solution, but what we are asking for is that we do our homework and do the study to provide us the information, so we can make an informed identity management solution.

Shaun Grannis – Regenstrief Institute – Director & Principal Investigator, ICEPHI
As I said in my testimony very quickly, the foundational theoretical principles of probabilistic algorithms have been well known. The seminal paper was published back in the late 1960's. Recently, some folks actually, I don't want to drag us into the weeds here, but recently some mathematical proofs about the … nature of this algorithm have been established, which have opened up new lines of research. So while the theoretical underpinnings are well known, the application and evolution of those algorithms absolutely are unpublished today. There are no transparent, peer reviewed, rigorous, scientific studies into what works in what circumstance.

I've been sitting on a paper for a couple of years now where some of the unstated approaches to how you use these algorithms, there's suggestions and hints and white papers or in conference proceedings. We've shown and we need to publish this that a particular approach to using this well-known algorithm does not work for newborn screening, for infant data because infants lack the first name. They lack social security numbers, and other parts of their demographics are missing. We've developed an approach that gets around that problem. I haven't published it yet, but nobody else has published it either. The best practices in this space absolutely need to get out on the table.

In full disclaimer, I am funded for a researcher initiated grant through AHRQ to actually look at some of these very specific issues. We are documenting the Regenstrief deterministic algorithm. I do research on probabilistic approaches, and we're going to be comparing them and publishing how the two compare among different cohorts and different types of data, different types of sources, so thank you for the question.

Paul Egerman – Software Entrepreneur
Thank you for the answer. Brad, you also have something?

Brad Malin – American Medical Informatics Association (AMIA)
Yes. I completely agree that the algorithms and all of their variants, as well as the additional rules that people append onto these probabilistic matching algorithms, are not very well documented. However, at the same time, even if you do have well documented approaches, you're not going to have complete clarity with respect to their transferability from one institution to another. Because we don't have any gold standard data sets that says, ―This is how you would apply this in Tennessee," or, ―This is how you would apply this in Indiana," because the populations are completely different, and their transience is different.

We also know that different age groups end up influencing data and the linkage in different ways. I think the Federal Trade Commission has shown this with respect to identity theft. People within the age groups of 18 to 24 are significantly transient, and so you have to adopt your algorithms to account for this semantic bias that exists within populations. So having some type of test data sets that cover the gamut of the problems that you'll run into in the world is extremely important. But this is extremely challenging because we're asking for identifiers, and so you're trying to create a data set that potentially is a privacy violation unto itself. So how you actually create such data sets and who would you allow to have access to them and set up that infrastructure for research is a long-term challenge.

Paul Egerman – Software Entrepreneur
Yes.

Garland Land – NAPHSIS – Executive Director
I'd like to comment that I think one of the previous speakers mentioned the fact that really the ability to link is dependent upon the quality of the data files that you're working with. And I think that's an important concept. We're involved in another project in which to link birth records with DMV records, motor vehicle records. We have purchased a software product for the states to use to identity the quality of their data because that's typically unknown. How many records don't have a name in it? How many records have a date of birth that is not a logical date of birth? So there are products out there now that can be useful to hospitals an providers to identify the amount of error in their data, at least to some extent. The problem is that once you've identified it, then you still have to figure out how you're going to clean it up. That's oftentimes a manual process.

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
I think it's an excellent question, and our written testimony does suggest that we have more research in sharing of those best practices. I would emphasize that in addition to work on the algorithms, we consider Wes' point about that human machine interface and that the research needs to include, in the trenches, what was the training. What were the policies? What was the management reinforcement? Were there rewards or disciplines associated? Was there feedback provided to those data entry folks? I think that element is such a big variable in this equation that we need to better study to understand that as well.

Deven McGraw – Center for Democracy & Technology – Director
Terrific. I have in the queue Carol on the phone, then Adam, then Joy, then Paul. Go ahead, Carol.

Carol Diamond – Markle Foundation – Managing Director Healthcare Program
I also want to thank everyone who offered testimony so far. It's been really great to hear and to hear some of the points that have been made. I wanted to, I guess, emphasize some of the points that have already been made, but also ask people to corroborate or not corroborate those points because I think, for us as a tiger team, it's very important that the takeaways are at the right level. I think everyone made this point that the inherent data quality of the basic information that's collected is at the heart of this issue. In other words, the better that is, the less and the more timely and the more quickly it's updated and all those good things, but the better that is, the less linking becomes the problem. So the cleaner, more accurate data that's collected at, if you will, intake, the less this challenge really is the challenge. That's one question or takeaway that I want people to comment on.

The second is that the solution, if you will, although we're not talking about solutions, but the linking approaches and the algorithms that get used are truly idiosyncratic, both to the purpose and the population that's being looked at. I say this because, in our own experience, particularly in Katrina Health, when we were working on some of those issues, some of the common and commercially used software to correct, for instance, spelling errors and common digit transpositions turned out not to be so good for French last names. There is a lot to be said for whether it's newborns or other idiosyncratic characteristics of a population that says that our goal is not the perfect algorithm or the perfectly studied algorithm. It is a level of accuracy, if you will. So I wonder if people could comment on those two takeaways because I'm trying to sort of listen across the testimony and come away with what's important for us to think about.

Barbara Demster – HIMSS – Chair, Patient Identity Integrity Workgroup & Consultant
What you've hit on is basically the heart of our white paper with the key influencers over the matching process. Algorithms are just one of those nine, and the data quality, data accuracy, all of the infrastructure that needs to take place to provide that good data to go into the system. Part of our operationalization of our recommendations is to develop a centralized resource for industry where we're trying to gather together all of the information that's out there because there is no central place to go. We have worked with the HIMSS privacy and security toolkit to establish a section where we will include industry guidance, documents, metrics, tools, sample monitoring reports, sample use cases, test scripts, all of those things, the guidance that's needed, as identified through those 56 recommendations so that we can start building an industry knowledge.

We're working with AHIMA, the American Health Information Management Association, and NAHAM, the National Association of Healthcare Access Management, to bring in their best practices. There's a lot of information that's already out there. It's just not available and accessible and known to the people that are in the trenches every day. So we have to address all of those business processes, education of the C-suite, the decision makers is incredibly important. Clearly, there's not an understanding of what is going on and what has to happen and the importance of data quality to their institutions. So there will be a section for important knowledge for that level of decision maker as well.

Shaun Grannis – Regenstrief Institute – Director & Principal Investigator, ICEPHI

Carol, I heard your first point. Better data leads to fewer challenges in linking. I wholeheartedly agree with that. A great book out right now by one of the preeminent statisticians at the Census Bureau, the Census Bureau knows a little bit about linking patients, and I suspect we'll hear about that in panel four perhaps, but they have a great book out that's called *Linkage and Data Quality*. These issues are not new issues. It's just becoming abundantly clear in healthcare how important these issues are. You can go back to Darrow, even Claude Shannon who came up with information theory talked about these issues back in the 40s and 50s. So we're just beginning to recognize that data quality is so critical to this.

In the research that I'm working on right now, it's abundantly clear to me that there are questions that we need to ask of data before it's matched. One of the dissemination of best practices should be what questions do we need to ask of the data before we link it? Some have been talked about, just completeness and accuracy and validation. There are many great commercial tools out there that address many of these issues. Again, the specific questions that we should ask of the data to understand its quality need to be performed.

Brad had made a point about what's the transferability of different algorithms or different methodologies. I would submit that no concrete, discrete, fully realized algorithm is transferrable. The process absolutely is. There's no doubt in my mind that there will be a probabilistic and a rule based component of any linkage algorithm that's created, and understanding the data quality, understanding the pitfalls in that data is crucial, so that's point one.

Point two; linkage approaches are idiosyncratic to the purpose and the population. I agree, but would want to clarify that again. I think there is a framework and a process for matching that many vendors have identified and work well, so the process, I don't think is idiosyncratic. The data is, and the tweaking and the tuning of that algorithm to different data sources is where the idiosyncrasies are, but recognizing the pathway that one must take to develop an appropriate matching algorithm for a given circumstance, I think there's a framework there that may be identifiable.

Brad Malin – American Medical Informatics Association (AMIA)

I would caution in one way, which is that people have a limited amount of time, a limited amount of money and a limited amount of bandwidth. If you spend all of your time trying to perfect the data coming in, you can limit your ability to, or limit the workflows that exist in healthcare environments. This doesn't mean that you should not do your best to try to standardize or clean the data. It's just that you should not necessarily invest all of your time and effort to make a perfect system because we'll never end up in a perfect environment. So this notion of having a little bit of dirty data or even more dirty data, you can end up with extremely dirty data, but have extremely accurate record linkage results provided that the noise that influences different records doesn't cause overlap in those records. You could have extremely noisy data and get good results. So while better data quality is fantastic, and we should always try to achieve it, we shouldn't try to force the system to over-specify.

Deven McGraw – Center for Democracy & Technology – Director

Moving down the queue, Adam Greene.

Adam Greene – Office of Civil Rights – Senior HIT & Privacy Specialist

The HIPAA privacy rule oftentimes requires covered entities to only collect and share the minimum necessary information, which may mean not sharing a date of birth or social security number.

Theoretically, I could see this being an obstacle for data linkage, but I want to get more information about in practice, whether in practice this has been found to be an impediment to data linkage.

Brad Malin – American Medical Informatics Association (AMIA)
Whether the absence of specific fields is an impediment to data linkage, is that your question?

Adam Greene – Office of Civil Rights – Senior HIT & Privacy Specialist
Whether covered entities attempts to comply with minimum necessary has stood as an obstacle.

Brad Malin – American Medical Informatics Association (AMIA)
The definition of minimum, if I may, can I respond with a question to say, is it fair to say that for a given matching accuracy, various fields may be minimally necessary to achieve that linkage accuracy. Is that a fair assertion?

Adam Greene – Office of Civil Rights – Senior HIT & Privacy Specialist
Certainly when looking at what may be necessary, data linkage is something that could be considered.

Brad Malin – American Medical Informatics Association (AMIA)
Right.

Adam Greene – Office of Civil Rights – Senior HIT & Privacy Specialist
And I don't want to stress HIPAA in particular, but as a general principle of privacy, there's oftentimes you should only collect the minimum necessary information, and how does that interface here?

Brad Malin – American Medical Informatics Association (AMIA)
Right. Yes, so with that, a little bit of extra information, we use as much information as is collected by the various systems for demographics to link the patients. So we do perform matching without social security number. We cannot perform matching without date of birth. Even if it's the date of birth is a little long, we can work with that, but—so age, but—so there are certain variables that are necessary in our algorithm. It may not be necessary in other algorithms, but we, for instance, see, generally speaking, a fairly uniform and consistent collection of social security number on about two-thirds of patients across our institution. It varies a little bit, and it's used. When it's there, we will use it to improve the accuracy of the matching process. I haven't experienced with any of our 80+ hospital systems, anyone saying we have social security number, but we're not going to share it for the purposes of matching. That issue hasn't come up in our exchange thus far.

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
I have not heard of a specific instance where the minimum necessary has impeded proper matching. However, in the written testimony, I did highlight the need to align various policies, and I think that's where there's a natural tension between the desire to have appropriate medical information shared, but also, of course, to ethically respect and legally respect patients' privacies. The best that we can do to align those policies to make sure that they are complementary, I think that's very important. I think you're bringing up an issue that can be a challenge where in order to maximize matching; there are challenges on the privacy side.

Barbara Demster – HIMSS – Chair, Patient Identity Integrity Workgroup & Consultant
It's an interesting question. I'm thinking of the information that's following after the match is where most people would focus there. Is it necessary or not? If you've already made that decision that you can share that information at some point in time, you would assume that the organization through its business processes would have had some level of understanding that that information was being passed through. But it also points out a need, again, a recommendation in our little white paper that we don't have any standards for minimum data matching criteria between systems. That would be incredibly helpful to have some level of understanding that you would at least have these data elements constructed in this format or whatever to help ease that understanding and process.

Deven McGraw – Center for Democracy & Technology – Director

Terrific. Thank you. Joy?

Joy Pritts – ONC – Chief Privacy Officer
I'd like to thank Adam for setting up my question, and Barbara for answering it in a way that segues so well into where I was going, which is, we had discussed a little earlier about the quality of data, and I am not a technology person, but I know enough to know garbage in, garbage out. One area that seems like it might be a potential area where it's problematic, and there may be relative, low hanging fruit is the collection of the data with respect to names, date of birth, things of that nature. How much of a problem is that in the matching process and how difficult would that be to address?

Brad Malin – American Medical Informatics Association (AMIA)
Can you clarify?

Joy Pritts – ONC – Chief Privacy Officer
For example, is it problematic if people record the date of birth differently from system-to-system, or the names, have the first name, last name; last name, first name. Are any of those problems or not?

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Typically, date of birth can be translated and normalized across systems.

Joy Pritts – ONC – Chief Privacy Officer
Would it make it easier if it was normalized up front?

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
I don't see that as a fundamental, one of the bigger challenges that we have because that date of birth, date conversions have been done within computer systems for so long. The main issue is an issue, and that has been brought up in several of the testimonies, special characters, hyphens, and middle names, and two last names, things like that, or some ethnic groups that may have the first and last name that are ordered differently than we're used to. Those types of problems are very real, and there aren't good standards for dealing with that, so I see that as a bigger problem.

Joy Pritts – ONC – Chief Privacy Officer
Do you think that you have, there's a comprehensive list of those types of issues with the names in particular?

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
I don't know that there's been a formal publication of those types, a comprehensive, formal, rigorous analysis of those types of issues. I can share with you a name, the process that we go through for name cleanup. I know that many systems rely heavily on a very long and comprehensive name standardization list, and I know that many systems invest heavily in identifying, for instance, ethnicity of names so that one can then correct any problems with that because name does become very important component with high informational content. It seems to me that there's been a lot of activity in investing in name cleanup. Are all of the problems declared? You can pick them out from different documents and different publications here and come up with a list of them. I think we, on the panel, could begin enumerating them, but I don't see that there's been a formal description of that.

Probably the best paper I can think of was by a gentleman by the name of Jerry Webber who wrote something back, a white paper on names and how they're dealt with. He worked for some company. I can't remember what was the name of it, but there's some good information out there, again, but it's not centralized. It sort of percolates up here and there.

Deven McGraw – Center for Democracy & Technology – Director
We'll go … everyone has had at least one chance to ….

Gayle Harrell – Florida – State Representative

Thank you so much. I appreciate it. Again, I am not a techno geek. I'm more of a policy geek, so I'm going to ask a question that's kind of off the wall perhaps, but in listening to you and understanding exactly what's going on, I foresee, as we get more exchange going on, the level of difficulty in liability issues growing. The more linkages you have, the more HIEs exchanging information across the country, the more difficult things become, and the more probability of incorrect linkages, false positives, things of that sort. I see great potential issue for liability. Can someone speak to that, and where does the buck stop?

Brad Malin – American Medical Informatics Association (AMIA)
So I'm going to spin your question. I'm actually not going to speak towards liability first. The first thing that I will speak towards is that actually the more data that you have, and the more locations that are actually sharing the data, you may actually end up with better accuracy because you may end up being able to resolve inconsistencies or things that would have been errors. One thing that actually does come out in my testimony is that if you have—and Shaun and I have had discussions about this. So if you have person A matching to person B, and person B matching to person C, but A and C are not linked, then you may actually have an inference there to suggest that A and C should be linked. So this notion of triangulation, of bringing people together in groups, you may actually end up with better accuracy in certain instances. So I didn't answer your liability question. I didn't answer the legal issue, but sometimes more data is actually more helpful.

Gayle Harrell – Florida – State Representative
I see some questions on other people's faces on that issue, but continue. I'd really like to ….

Brad Malin – American Medical Informatics Association (AMIA)
Do you want to take a crack, or I can take another one?

Shaun Grannis – Regenstrief Institute – Director & Principal Investigator, ICEPHI
Yes. I think, related to more data can lead to more accuracy; I think that is true if there's a high degree of coordination in the cleanup effort amongst those different parties. If there is less and less involvement along the chain, as it passes, and the endpoint of the data recipient is unaware of perhaps even the source and the degree of quality of the source, etc., and there's not much centralized control or cleanup, then it becomes more difficult. I think it partly depends upon if it's a health information exchange, for instance. If it's highly centralized and controlled, I think that would be true. If it's not, then I think the data quality could actually be worse. In terms of liability, I'm sorry; I'm not able to comment.

Brad Malin – American Medical Informatics Association (AMIA)
I think what you're hearing is that the premise that as things grow, the level of difficulty will grow. That may be true. It's not been proven, and that's why, in my testimony, I said I think, as we expand this, we need to study it and understand it. So we have state of Indiana. We have the state of Tennessee here. We've talked a little bit. Is there some way to do some testing across larger systems to see what happens in healthcare?

I know that we've looked at this in other spaces, but again, because data quality is so crucial, and the data quality in healthcare data is different from other industries, then we need to answer the question. It may turn out to be that it's just the same as other industries. We don't know the answer to that question today. So I think that is crucial to recognize that we need to answer that question.

There was another point I was going to make. In terms of liability, I think one of the other points in my written testimony that I didn't state verbally is ONC needs to help with messaging that matching systems will never be perfect. Never, and so this notion of liability is intimately related to what the public's expectation of the system is, so we need to set reasonable expectations for what the capabilities of the system might be. I think that may mitigate the liability issues somewhat.

Joy Pritts – ONC – Chief Privacy Officer

… since ONC was invoked in that last one, I have a follow-up question, please? I'll keep it short. Does anyone have any information about the accuracy of medical record keeping in the paper world for comparison sake?

Brad Malin – American Medical Informatics Association (AMIA)
I do not.

Shaun Grannis – Regenstrief Institute – Director & Principal Investigator, ICEPHI
Yes. I've not studied that. No.

Barbara Demster – HIMSS – Chair, Patient Identity Integrity Workgroup & Consultant
I don't, but I'm going to research that, and I'll get back to you.

Deven McGraw – Center for Democracy & Technology – Director
Paul?

Paul Egerman – Software Entrepreneur
Actually, I first was going to comment on the question that Joy just asked about the paper world. It's probably hard to get data about that for lots of different reasons, but I would just tell you my own personal experience is there are a huge number of problems. Was it the PCAP, the Webinar that occurred yesterday, Larry Summers talked about a situation where he had a personal situation where fundamentally there was erroneous healthcare information given because somebody had handwritten a name. Instead of Summers, it was really Simmons, and there are a lot of issues in the paper world, and it's a really excellent question because nobody should come to the conclusion, in my opinion that somehow computer systems are creating a problem that didn't already exist. I would tell you, back in the early _70s … industry working at Mass General Hospital, and a physician in charge … Barnett, had teams review paper records and report on them. Our team physician had a sense of humor and reported that the patient in 1967 had basically an onset of diabetes, which is accompanied by an odd change in race, and the medication worked extremely well, and in subsequent visits no longer occurred. That problem no longer occurred. Obviously … the encounter was put, the wrong encounter was put in the record because the records back then were very difficult.

But getting back to the question I had, I also want to thank the panelists. Excellent presentations and a great deal of work went into these. My question is, it was very interesting, first to Scott Whyte. You introduced the concept of false positive and false negative, which is in a lot of the other presentations also: false positive being when you've got the wrong patient information or multiple patients in the record, and a false negative being when you have duplicates. I just want to make sure. I'm under the impression that in terms of the numbers that duplicates is the far bigger problem. That because people want to avoid the false positives, we end up with a large number of duplicates. In your written testimony, you said you have a duplicate rate of 8% to 12%. Is that impression correct?

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Yes, that is correct.

Paul Egerman – Software Entrepreneur
And the duplicate problem is a significant patient safety issue also, right? I mean, it's ….

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Yes, because a clinician would be missing the information that may be important to the treatment of hat patient. There may be allergies or medications or other diagnoses that the clinician is not presented with.

Paul Egerman – Software Entrepreneur
Okay. Now if we could turn to the false positive side for a minute, if I look at that broadly and say, what are the reasons why there might be the wrong patient's data, for example in my medical record, some other patient's data in my medical record? All of the testimony that we're getting sort of points to various linking problems. But isn't there also another source for that, which could simply be either poor clinical

processes or simply human error? Basically, a laboratory test result of a patient, somebody draws blood from me, but enters the data into somebody else's record, perhaps because they're busy, they're stressed out, or they don't have the correct processes in place. Isn't that another source for this kind of an issue?

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Yes. Definitely, there are clearly people issues. There are process issues, and there could be technology issues, bar-coding, or there are other tools that can improve the accuracy of that data input, but there's still people involved. Sometimes it's an error on the part of the patient, so the patient may be sharing identification or things or may just give incorrect information to someone who is registering, and that causes the data to be inaccurately captured.

Paul Egerman – Software Entrepreneur
Thanks. That's helpful. You also said that the size of the organization has an impact on the accuracy. The sense I got is the larger the organization—if I understand your written testimony—the larger the organization, the less connected the physicians and clinicians were to the patient, the more increased likelihood that this problem would occur, that there'd be linkages problems in general?

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Yes. I have observed informally that correlation. You think about a physician office where they might know your name and face. They're typically going to collect your information correctly because they know you, and so it's the smaller communities that are typically the case, and the same can be true for very small hospitals. Midsized hospitals, people are more anonymous. That can be difficult. Sometimes they also, however, don't have the economy of scale to have infrastructure of the more expensive tools or additional labor to correct the problems. There are some instances where larger organizations, research organizations, academic medical centers, for instance, may have more resources to put on the problem, even though they're larger.

Paul Egerman – Software Entrepreneur
So if you think about it from the standpoint of interoperability, if a clinical summary is transmitted from a small medical group, from a two- or three-person partnership, the sense I have is the greater sense of confidence that that—in the quality of that data compared to a much larger entity where there might be some possibility that you're getting—that duplicates exist. Maybe even a 10% probability that duplicates exist, and you're not getting all the data.

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Yes. I think that's reasonable. I don't know that it's been studies. But, yes, that would seem to be a reasonable conclusion. It also depends on how transient the patient base is. If that medical group is a primary care physician, then the likelihood that data is going to be good is probably true. If it's a community clinic that may not have long-term relationships with a patient, even though they're small, they could have less accuracy.

Paul Egerman – Software Entrepreneur
That's helpful. Then I don't know if there's— I actually had a question for Adam. It was interesting in Garland's comment where he said he was linking birth records to a motor vehicle database for, I guess, research purposes. I was curious about that from a privacy standpoint because I feel, when I apply for my driver's license, I didn't know that that would be involved with my healthcare delivery. Does HIPAA or are there any privacy issues that relate to motor vehicle databases? Is this allowed to do?

Garland Land – NAPHSIS – Executive Director
I can explain that. It's really not for research purposes. There's a federal law that will be coming into effect in 2011 that requires motor vehicle agencies to, when you present your birth certificate to get a driver's license, they need to verify that your birth certificate is indeed a valid birth certificate, and it's not a falsified document. So we provide a software product for the motor vehicle agencies to use to enter key information from the birth certificate that then pings against the vital records database to insure that that

vital record is on file. That's why I was referencing the fact that it's important to have good data in the birth certificate file for that matching to occur … law.

Paul Egerman – Software Entrepreneur
That's helpful to understand, Garland, but what I'm still curious about is when people are talking about patient identification systems, and especially look forward to the idea where everybody – we have universal coverage. Can we link to driver's license databases and use that as a vehicle to help us with this whole issue?

Garland Land – NAPHSIS – Executive Director
I think it varies by state. In some states, motor vehicle files are public information, and they are purchased by various corporate entities for these types of matching purposes, and in other states, the motor vehicle files are private and cannot be obtained.

Brad Malin – American Medical Informatics Association (AMIA)
… it's not just DMV records, so there's all different types of publicly available resources, but it also depends on states. For instance, voter registration databases are also used for verification purposes, and we, in particular, use state voter registration databases for testing our record linkage algorithms, but you can't do that with data from the state of Tennessee per se because you're only allowed to use it for political purposes. But we can take data from the state of North Carolina and use that for evaluation purposes.

Paul Egerman – Software Entrepreneur
Here's my question. Is there any federal impediment to doing that?

Brad Malin – American Medical Informatics Association (AMIA)
No, not ….

Adam Greene – Office of Civil Rights – Senior HIT & Privacy Specialist
For federal databases, you generally have the Federal Privacy Act. Oftentimes states will have privacy acts governing their databases, such as DMV databases, and there can be various significant differences between the states on how that information can be used or shared.

Deven McGraw – Center for Democracy & Technology – Director
Great. Thank you very much. All right. We have some requests for a second round from three folks, and we are going to end this panel on time, so I'll ask each of you to try to be as brief as possible because we only have about five minutes left. John?

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
To follow up on something Shaun had said earlier about the situation where you have a lab data being sent to a provider, and they have enough information in order to link it. I guess the question is have is, I'm sure a lot of people will say that data is good enough for me. How do you deal with that paradigm, which is, it's good enough for me for the purpose I intended it for. It isn't wrong. It's good enough for me. The same lab maybe in the future or a pharmacy may need to provide data to another provider because it's now part of a larger rollup. It sort of speaks to the need, in my mind, to either have some set of standards that really describe what is necessary demographically in order to insure that we can match patient information. I'd just sort of be interested in understanding what your thoughts are on this and then who is responsible when somebody comes back and says, it's good enough for me. I don't think I need to change it. I don't have a data quality issue.

Brad Malin – American Medical Informatics Association (AMIA)
In the interest of time, just some very quick responses: For a data source to join our information exchange, there are minimum fields required, and if the source does not provide those minimum necessary fields, they can't participate in the system. There are different types of sources. There are authoritative identity sources that come into our system, and then there are results sources that can piggyback on top of authoritative sources, so registration systems are authoritative identity sources. If a

lab gives us a local medical record number and very minimal other identifying information, we may be able to link that result into the system, but not declare or modify an identity based on that. We have, again, there are details here that would need to be gone through, but we have processes in place to do that and minimum standards for participation in the system.

Deven McGraw – Center for Democracy & Technology – Director
Wes and then David.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
If there's one thing that struck me here today, it is the statement that we don't even know how to ask the question about how good the data is. I think that a takeaway from me is that we need to find ways to characterize the accuracy of the data. I was particularly taken by Shaun's example of two different use cases for matching data that had radically different decisions about whether to emphasize false positives or false negatives. So I would say that we shouldn't just turn this into a question of absolutes. We should turn it into a series of questions, how do you characterize the quality of the data that's the input to the matches. What is the quality based on that characterization? What can you tell about the quality of the matches when you do that? To the standard way of looking at things, which is garbage in, garbage out, I would add there's also garbage in, biofuel out, so we need to think about the quality that we're looking for.

Deven McGraw – Center for Democracy & Technology – Director
Thanks. Okay David.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Mine is a comment rather than a question, just to register with respect to the discussion that we had earlier about algorithms. When I was using that term, I was referring to what I think, Shaun, you called it framework, which is really the notion of an algorithm that is broad enough to adapt itself to the local variations and can be tuned appropriately to reflect the prior probabilities of data that could throw the hard coded algorithm off. So perhaps we need to be more specific when we use words like algorithm, but I'm referring to a systematic approach that could be applied to any population, perhaps the machine learning or other localization capabilities—just a comment.

Deven McGraw – Center for Democracy & Technology – Director
Terrific. You did just as I asked, and we're ending the panel on time. Thank you very much for sharing your testimony with us and for sitting and answering our questions. It really kicked off the hearing in a very good way, and I know I for one learned a lot, and I suspect that's true of our other tiger team members as well, so thank you very much again, and you all are either free to go or stay if you want to hear the rest of the day. Thanks again. We'll take a five-minute break while we assemble the second panel, so don't go far, but you've got five minutes.

All right, so before Paul introduces the panel, we're switching roles, and I'm the time task master, and he set a very high bar, and so in fairness to the other panel that got pretty, where we adhered very tightly to the five-minute time limit, we're going to do that. But as you notice, there's a nice question period afterwards, so if there are points that you don't get to emphasize, you will more than likely have an opportunity to do that during the question period.

Paul Egerman – Software Entrepreneur
This is the second panel. We actually have two panels on what are the solutions, and so it's not necessarily anything specific about how we broke this up, but what we're interested in hearing is how people are addressing the whole issue related to patient linking. So the first four presenters, we have an excellent group. When I read the people's biography, I always feel badly about trying to introduce people because I figure I can never get anywhere close to what your actual accomplishments are. So what I'm going to do is ask you to introduce yourselves simply by saying a sentence of who you are and what you do very briefly because I feel your own modesty will keep it brief, and that why I'm off the hook from not having accurately described you. I'd ask you to first briefly introduce yourself, and then we'll get started.

Why don't we start with you, Rich?

Richard Elmore – Allscripts – Vice President of Strategic Initiatives
Good morning. I'm Richard Elmore. I'm with Allscripts, Vice President for Strategic Initiatives.

Scott Schumacher – IBM Software Group – Chief Scientist
Hello. I'm Scott Schumacher with Initiate, now part of IBM, and I am chief scientist.

Paul Oates – Cigna – Senior Enterprise Architect
Hello. Paul Oates with Cigna, a national health plan, a senior enterprise architect is my day job, and my hobby job is chair of the Patient Centered Payer Round Table and other committees within HIMSS and other workgroups.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
I'm Sara Temlitz. I work for the Department of Veterans Affairs. I'm a business product manager within the data quality program.

Paul Egerman – Software Entrepreneur
Terrific. Thank you very much. I think, Sara, you are the first presenter.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
Good morning, everyone, and thank you again for inviting us to provide this testimony for this very important topic. The timeframe works perfectly for me because I only have brief notes for my introduction, and then I have written testimony that I'm hoping we're going to cover in the question period, so I just wanted to give a little background about the VA itself. We do have 153 or so medical centers. We do have a centralized master veteran index that holds over 16 million unique patients. We've been supporting that for over 10 years now, which has built us up to that 16 million plus patients. We also have a staff of over 40 data analysts within the data quality program that is designed to be the infrastructure that really supports the patient matching and all the activities around that activity.

I wanted to note that in the testimony today and in the questions that we answer, generally I'll be responding from the internal perspective within VA. We also do have activities that we're working on with interoperability with external partners, other federal agencies within the National Health Information Network, so if I'm speaking for that interoperability, I'll specifically say that.

Some of the important points that I wanted to cover that are in the written testimony are, we really strive to have a robust set of standardized traits. We feel it's really important to adhere to any standards that exist around the process of patient matching. From the trait standardization, what traits we use, if there are any standards for what kind of format they should adhere to. We also use the object management group or OMB patient identifier service standards, ASPME 1714, so wherever there are standards that do exist, we adhere to those and have implemented those in what we do.

We also have put a lot of focus on; there's been a lot of talk about the algorithms, probabilistic, deterministic. We spent several years really looking at how our matching works within our patient population and what worked best for us, so we spent some time doing research on how we can improve our algorithms and the other processes around that. We've also built some tools that re used by this group of data analysts that we have and focused a lot on what we need to make sure that we're maintaining those linkages, that they're happening the way that they should be, and that any of the support around the matching itself is supported by the toolsets.

We also have developed some strong internal policies and support for data quality and the resolution of any identity matching issues, so we think it's important that you look at the full spectrum of what patient matching really means. It's not just matching records upfront, but also maintenance of making sure that you don't have errors, how do you resolve those, and how do you tighten up the actual matching process so that it's working as best that it can?

The other thing that I wanted to cover really briefly were some of the lessons that we learned. It was one of kind of the general questions for the group, a lot of talk about data quality and how is it important. Well, in my introduction, I said that I work in a data quality program. Really, our patient matching and identity management service stems from our data quality program. It's where it started. We really focused on the identity traits, defining what we felt that set of traits was going to be, working on the data quality of those so that we had a good basis to start when we were doing our linkages for patients.

So we feel that that's really still important. We still put some deal of focus on that. We do some data analysis and data profiling on those traits to make sure that they are still at the level that they should be.

We also have developed business rules for updates to identity data. We feel that it's important that we have really the best information that's shared with other groups. We realize that some medical centers may use nickname, and some may use full name in our record of what we call our primary view or the gold copy. We always want to have the best information there.

We also have looked at thresholds for defining matches, potential matches, non-matches. We're talking about the false positives, the false negatives, and how important it is within healthcare and the longitudinal health record that we do err on the side of caution, and that we aren't matching any patients to patients that shouldn't be done so that we can maintain the integrity of those records. Along with that, we have processes and tools to detect when there are errors, catastrophic edits, we call them, when we may have mixing of patient record information, and how we can resolve those and prevent those from happening. Also, our infrastructure of tools and staff that I mentioned earlier, including training, and looking at really learning lessons of what we've done and how we can continue to improve and tie that all up. Thank you.

Paul Egerman – Software Entrepreneur
Thank you, Sara. This is a great presentation. I'm always impressed with the work the VA does. That was really excellent. Next, we have Paul Oates from Cigna.

Paul Oates – Cigna – Senior Enterprise Architect
Hello. I'm very appreciative of the opportunity to be here. For the last 20 years, I've been spending time as a health plan with the collaboration topics between payers, health plans, and providers, and often get asked why someone who is wearing a black hat is primarily concerned with claim payment is in this topic and these issues. This is our opportunity to join the team. We care a lot about that meaningful use stuff. And if there's an initiation process that lets us sit at that table, we're all for it. So I'll take off the black hat and talk a little bit about healthcare and health plan roles that are changing.

Known for most of the time around financial insurance processes, we're now becoming engines … improvement and offering the informatics and the meaningful use kinds of wellness processes needed to help power that topic. We care a lot about the identity issue and have the same exact challenges that the healthcare professions and hospital systems had. The way we're solving those is to use probabilistic matching tools, so to specifically answer the question, what are the solutions? We're using those tools, and we are linking information on a national basis across both administrative, traditional eligibility kinds of data, and clinical data: lab, pharmacy, biometric results, and seeing some success. We can attest to the scalability, accuracy, and value of these tools.

We use them. They're not perfect, but they do work, and my colleague to my right, Dr. Schumacher, can give you the details on that. So we're solving overlaps in data sets of between 6% and 37%. Those numbers are in our context. They really can't be applied to any other organization because it's very contextual. We don't find the need for different solutions for real time or dynamic linkage and batch or informatics linkage. In fact, we think it's pretty important to use the same solutions for both because it does little good to probabilistic, do the sweat and hard work probabilistically to match up the data only to have someone deterministically then try to consume it, so hence the need for the right architecture to wrap around these solutions.

The other thing I really wanted to bring to the table, if I could take a little bit of liberty, is to try to broaden the discussion. The patient linking hearing here is largely focused on linking of data about people who are flowing through hospital and provider systems and an HIE, as they're treated for an illness. I think rather than patient identity linkage, a better context might be consumer identity linkage. The reason for that is that when an individual is playing the role of a consumer trying to actively manage their care or when information is flowing between consumer, provider, and health plan, say an accountable care model, then the identity problem extends to all the other places where identity flows.

So the question was asked whether consumers are concerned about this. Do they see the impacts? They don't see the impacts behind the curtain, but they see it when they can't navigate and they're frustrated about being able to get the information to make decisions.

So it's entirely possible that our healthcare system will not be able to successfully implement accountable care organizations, medical home, or other collaborative care models unless person identity is linked across the entire ecosystem, and that's what I think we're all trying to accomplish. A simple example of this is linking, for example, stage one meaningful use core measures gathered in a physician's office like wait, to the availability of a healthy eating program at a health plan. We're working very hard to do that today.

What should ONC do? The first would be to recommend broadening the discussion. Include wellness, ACO miles, and other processes that health plans and others do in the discussion. There's not a full-scale model of identity flows across the entire healthcare system. I don't know of one. We need to develop and maybe task a workgroup to put the big Where's Waldo flowchart on the wall and look at where all these linkage occurrences happen across the entire ecosystem. Until we have that, we really don't know where the problem is.

The other thing I'd like to reinforce in my last 45 seconds is the data quality issue. We've spent more money on data quality than we did on the matching tools. I think you might want to have another hearing on quality. It's more important. We can probably achieve more utility in the matching discussion by investing not in a universal patient identifier. If we wanted to do that and we can get it, fine, we'll use it. It's helpful, but not sufficient. But instead, in best practices and training and techniques for data quality, and perhaps the regional extension centers and others ought to be training on data stewardship, master data management, and data quality kinds of topics because that is the secret sauce to this problem.

Paul Egerman – Software Entrepreneur
Excellent. Thank you very much. Scott Schumacher.

Scott Schumacher – IBM Software Group – Chief Scientist
Thank you very much. I probably should wear a black hat as well because I'm an evil vendor in this organization. I'm with IBM Initiate, and we have been actually making a living doing probabilistic matching solutions for over a decade now. We started within healthcare, but we actually do apply the same technology in insurance, in retail, law enforcement, banking, and other industries. It's a very, very common, well-understood technology. In fact, about 85% of the citizens in the United States and Canada are being matched by our technology somewhere for one of our customers. Many of our customers are on the panel today.

I want to echo what Shaun had said earlier that really the probabilistic approach or matching, the technology behind it, the science behind it is really well understood and mature. So what I want to talk about a little today is emphasize some issues that I've put in my written testimony is what is accuracy. What do you do to get it? What's the process we go through to install one of these systems? What does it mean? I think, when we get down to it, the issue is, what accuracy is required to make a system work, and how do you achieve that? I'll echo a lot of the comments from the earlier group.

The good news in this business is there are only two things that can go wrong, a false positive and a false negative. The bad news is it will go wrong. There's no perfect system, as everybody has said before. If you're going to play in this game, and you demand no false positives, then you will have a very, very high

false negative. So understanding and determining what's an acceptable amount of errors on each side in order to make the system work is a key piece.

At Initiate IBM, what we do is when we go in and start this process with a customer is we first get in the first issue, which is, what's an acceptable false positive rate? That's, I don't care if you're building a missile detection systems or patient matching systems. The first question you ask is what's your acceptable type one error? Once you have that, that tells you the likelihood or the probability you need to accumulate in order to match a record. So if you say typical hospital environments, most of our customers will run on something like one in a million false positives. That is a million queries should give you a false positive. That's a typical rate you would go to do that one.

Some you'll hear later require a little higher one, like one in a billion. These are all achievable. But what you need to do is specify it out front. For your particular application, what's an acceptable false positive rate?

Once you have that, matching, as we have said throughout this, is an information problem. How much information is on the record, and that's what goes to determine your false negative rate. So once I've decided I need a threshold and likelihood of 12 in order to get my false positive rate, how do I get that likelihood? That's where you come into data quality and data completeness, and that drives your false negative rate. False negative rate, if you go to a hospital, you heard from Scott Whyte and others, the typical number you're going to see in a hospital is something like 8% to 10% if you're not doing matching, and it's going to be down around 5% to 2% if you are doing matching. Those are the numbers you would expect. Those are achievable within typical data quality issues.

Let me separate for a moment data quality from data completeness. Data quality problems, I've put typos in the names, I've used French, I've got French last names, and I don't understand phonetics on those. Those issues, the algorithms take care of, and there's been a lot of work over the last decade. In fact, that's the mathematician full employment act is, as we have gone to new data and more data, understanding the algorithm issues we need to take care of data quality issues. You cannot compensate for data completeness. If data are missing or invalid, that is going to contribute to your false negatives.

When we start this, the process you go through with doing this is to now profile the data that's available, determine the … the information across those attributes you've already used for linking, and then determine from that what you can achieve in terms of a false negative … system. I think what I would suggest to this committee is that you not establish a minimum set of attributes, or you not say how you're going to match or what algorithms you're going to match. But say for each particular cases, what errors are acceptable. I want a false positive rate of one in a million. I want a false negative rate of 10%, and then let the consumers of the people have this figure out what data they need to do it if they have it. Do they have to put eyeballs on it? How do they go to achieve those rates? Thank you.

Paul Egerman – Software Entrepreneur
Excellent. Rich?

Richard Elmore – Allscripts – Vice President of Strategic Initiatives
Good morning. Thank you for the opportunity to participate today. Allscripts provides clinical solutions and services to a broad base of healthcare stakeholders, including 1,500 hospitals and about one in three physicians in the U.S., and we're deeply involved in building a connected community of health on this foundation. As the distance between settings of care grows smaller, and as there is more interconnectedness between providers, the importance of accurate patient matching increases. As you've heard, secure technology exists for accurate matching.

Scott emphasized that the key is the data behind the match. Source data quality, completeness, and consistency must be addressed. It's also important to be reminded that there are large databases out there that have historically dirty data in them that is not part of active matching, but is the foundation on which new matching is taking place, so it's important to keep in mind as we move forward.

As an example, so source data quality is by far and away the most often reported route cause of patient matching problems. As an example, Steven Anderman at Bronx Lebanon Hospital, COO, has led major advances in care coordination and health information exchange. He has the technology, but Bronx Lebanon Hospital's patient population is two-thirds Medicaid. That addresses some shared identifying information occurred frequently. So an undue burden is placed on the provider to be the regulator of patient identity. The patient-matching attempt repeats itself at each healthcare organization in the community, and Bronx Lebanon is not alone. As Scott Whyte emphasized in describing his testimony, Catholic Healthcare West has had to place some major emphasis on people, policies, executive sponsorship, and monitoring to improve data quality.

Patients are not the only constituents. Guarantors and subscribers are typically difficult to match. The tiger team should consider individual person matching more broadly, which can be addressed by a robust, statistical and probabilistic matching like Scott and Paul have described. Patient registries for health information exchange will, of necessity, be interfacing with large and small care delivery organizations, and all must be able to participate … 70% of healthcare is provided in small practices, and these providers must have access to the same levels of capability, performance, and supporting tools for patient matching as larger organizations.

Much of the innovation around patient matching already has its … in advanced healthcare organizations like Sharp Healthcare where Bill Spooner's IT team supports 7 hospitals and 2,500 physicians. In 2009, on a base of 4 million patients, they identified and managed 260 manual registration errors, not a very small number. In the same timeframe, 2,740 duplicates were generated off of enrollment tapes. As more of healthcare becomes risk bearing, there is clearly a lesson to be learned around Sharp's experience with duplicates and enrollment. Bottom line is ONC should look to the more experienced healthcare organizations for continued innovation. It isn't about the match as much as it is about the quality, consistency, resilience, and recovery capabilities around the match. These innovations are emerging at places like Hartford Healthcare System where Steve O'Neil's IT strategy calls for leveraging the patient registry and the HIE infrastructure to create a patient throughput platform for heterogeneous provider systems. This requires robust technologies and processes for patient matching.

One solution that isn't likely is that third wheel of patient identity, the national patient identifier. However, the industry should pursue a unique, voluntary identifier, as it will improve matching for those patients who opt in with a key word, of course, being voluntary. It's not, however, the silver bullet, as other testimonies indicated, as it would take time to implement, and there will always be a need for patient matching technologies.

In summary, solutions for the tiger team to consider should include, first, insuring high quality demographic information using workflow, training, strong attention to process, and EHR and NPI technology with resilience and recovery built in; also, adopting the recommendation of the HIMSS patient identity integrity workgroup that Barbara Demster presented this morning. Second, establishing performance standards for patient matching algorithms, that's much to the point that Scott just raised about establishing the standards for performance rather than the specifics of the algorithm itself. Third, standardizing identity technology around the match for health information exchange, including IHE profiles related to patient identity, including PICs, PDQ, pediatric, and cross-gateway patient discovery.

Also encouraging best practices for governance of data quality, which occur in different models in the distributed environment, and just a brief note that there are emerging versions of the profiles I mentioned that fit nicely with the health IT report that was released yesterday by PCAS. Fourth, standardizing device communication, and lastly, encouraging patient engagement capabilities like those you're going to hear about from Sean Nolan in the next session that can improve demographic accuracy through patient verification. In conclusion, accurate patient identification is fundamental to healthcare quality, efficiency, and safety, especially as EHRs and HIEs become more embedded in the healthcare system. Thank you.

Paul Egerman – Software Entrepreneur
Thank you very much. Thank you, great presentations, very interesting from different perspectives. We have Sara who is a healthcare provider. We have Paul Oates, a payer, and we have two vendors.

Before we open it up to the team members who have a number of questions, I actually have a question. I want to pick up on the comments that you made, Scott, where you suggested that there be basically a guidance or a number given about what is the accuracy rate that should be expected. I'm curious to hear from the other panelists their reaction. To rephrase the question, should ONC establish a rate? Should ONC, for example, say 10% is the number that you should have for false negatives or duplicates in your system? Would that be beneficial? Is that the right thing for ONC, for the federal government to be doing? Any of you can talk … some example.

Paul Oates – Cigna – Senior Enterprise Architect
Sure. I think it's advisable to start putting some of those best practices in place. I'd have to think further about actually setting a standard because it's so contextual. There are different use cases that would drive different false positive rates. A highly intense patient care situation probably deserves a more accurate false positive rate than maybe an informatics or analytical process.

I think what would also be helpful is to understand for a given rate what are the implications that an organization would have to consider to meet a rate. Scott is right. You have to set your expectations about what your rate is, but then identify what it's going to take to get there in data stewards, staff, process, quality improvements, and I think that's really the discussion point is what's the balance between those two, the resources needed to get to the rate versus the rate itself.

Scott Schumacher – IBM Software Group – Chief Scientist
I think that the absolute number is probably not as critical is that there be one. That by putting out a number there, what you're doing is you're saying to everybody, you need to measure your false negative rate. You need to measure your false positive rate. We heard in a lot of the testimony today that people aren't quite sure what that is. They don't know what it is, and certainly in a paper world, and many people in an electronic world don't know that either. So the point of establishing a number is that people start to develop processes, and they are processes.

It's not just easy—I solve an integral equation and now I know the answer. It is a process around the data and understanding the data and doing the manual review that's got to be an issue. But I think the key thing is to put the processes in place to measure the errors, not just to say if you dictate the process, what you're going to get is you're going to get errors that vary depending upon location. Matching, and Sean and I argue over dinner, had a wonderful time, about if you're matching in areas with large ethnic populations, the same algorithm is going to give you different answers. If you specify what the requirements are, people are going to measure them and go to that without telling people, we'll measure it. My fear is, no one will look at the errors or the error rates unless you put a requirement on it.

Richard Elmore – Allscripts – Vice President of Strategic Initiatives
The Rand study in 2008 in evaluating this basically came up with about an 8% error rate that was tied to gaps in the data that was available for the match. So perhaps another consideration for the tiger team is the evaluation or the measurement of data quality or data accuracy.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
I agree with what they're saying, and I also agree with what Paul said in his initial testimony that I think master data management, data stewardship, having that structure within your organization is important so that you can look at these issues. I agree that saying a magic number is what makes us all hesitant to even come forward and talk about it is that I don't know that any of us want someone to tell us what that number is because being in healthcare; I said to Scott, our duplication rate is like 3% to 5%. That 8% to 12% made me a little edgy. So I think that looking in perspective that we really need to, as organizations, do the analysis and know that we need to do that so that we come to the right conclusions.

Paul Egerman – Software Entrepreneur
That's very helpful. The role that ONC might be … encourage or make it possible for people to do the measurements, but as opposed to set any specific target, which suggests each institution will have to set themselves. Very helpful, so let's start with you, David.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
I want to maybe broaden the question or the space of our considerations a little bit based on the expertise on the panel and the experience that you guys collectively have to include the question of authentication. So we've talked a lot about data quality, that if the data isn't captured accurately or completely, the good algorithms will be relatively useless because there's no data to drive them. What have you learned, or what are your thoughts about validating that the data that's being captured is in fact accurately belonging to the person who is presenting the data? For example, the use of a picture ID to validate that the person who says my name is John Doe matches the picture of John Doe? Or questions from a knowledge base to challenge the person at the time of presentation of their data, is there a role for that in what we're talking about?

Richard Elmore – Allscripts – Vice President of Strategic Initiatives
Yes, there absolutely is, David. The source data quality is vitally important. A number of the providers that we consulted in preparation for this testimony indicated that they have processes in place, number one, to secure the correct information upfront from the patient when presenting, the provider themselves being, their caregiver themselves being a last result verification of the patient's identity, and really trying to take the care at the data source. That's number one.

Secondly is to the question about photos. We do have customers who are using multifactor lookups, both in terms of photos and biometrics. Palm vein scanning is technologies that can be used, which can improve that. As you begin to think about health information exchange, I believe that something like biometrics isn't into today's IHE profile, so that there would be some challenges in that regard. But certainly these are techniques that are being used to get the specificity that allows for accurate patient identification.

Scott Schumacher – IBM Software Group – Chief Scientist
I think it's a great point to bring in a knowledge system, and that's one of the things that other industries think about when they're talking about matching or identification of consumers is that there may be some national knowledge source that can be used. That's an interesting thing. Certainly, for example, when you're matching organizations. You think about using Dunn and Bradstreet or Info USA or other sources of truth. I think knowledge base is an interesting question and should probably be explored.

I think the second piece of that in terms of authentication, in terms of pictures and biometrics, I think that that's obviously the future is that there will be more complicated, high value attributes. I think about a thumbprint, or I think about a picture, or a voiceprint as an attribute like name, etc. It just is a very value, like an identifier. I think, as we go forward, that that will be key to identification. We have some customers, not in healthcare, that are combining biometric and biographic matching in a single solution, and so it exists. The difficulty is that I've got about several billion healthcare records out there, none of which have any of this information on that, so this is really a go forward approach for biometrics. However, one of the other options is that there is a lot of in fact other information on healthcare records, blood type, race, etc., weight, height that could be used as part of an authentication challenge response system, as you talked about.

Paul Oates – Cigna – Senior Enterprise Architect
I talked a little bit about the need to broaden the discussion, and I think the comments from Rich and Scott are accurate and very relevant to patient care situations where the patient or the person is standing in front of you. What we're starting to realize is that a greater number of interactions will occur when the person is not standing in front of you, whether it's a person trying to access and register for their personal health record, or to schedule an appointment or do lots of online work where we can't see them. So there is a tremendous amount of information that's available in the public domain that's gathered about individuals, and there are companies who are organizing that information and creating knowledge based authentication solutions. We're starting to implement those.

So it's possible to query someone and ask them whether, in 1972, they owned a Dodge Dart of a Valliant in Las Vegas because they're the only ones that are going to know that, and then by answering that question, you know who they are. This is kind of a silly example, but you get the point that those

knowledge based authentication solutions can offer verification services at the point of interaction when you can't see the person. So when we are implementing those and starting to do that now, we're adding that information to the repository of data about the person so that we can use it in a go forward basis interactions with them.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
It's interesting to hear about solutions that are in the works and that are future based. Within VA now, we focus on a pretty rudimentary, asking for a photo ID. We have policies in place. We have training that we do. The VA does issue their own photo ID card for veterans that, in that process of issuing that ID, they check more than one. They have a cadre of different pieces of documentation that they can bring in, in order to get their card. We do encourage them to use their card. There is some focus on that front end, making sure that the patient really is who they say they are.

We are venturing into some of the different areas for online authentication. Specifically in areas where we're doing some interoperability, say, with the Department of Defense; the transition from active duty to veteran status, where they may have some IDs that exist, and which ones do we accept, and what information was used to create those so that we can insure that that information really is accurate. So I think that definitely there are some areas where we will be moving into, but right now a lot of focus is still on that person-to-person interaction.

Richard Elmore – Allscripts – Vice President of Strategic Initiatives
I'd like to add that there's the potential, based on the feedback that we've gotten from a number of providers, for real significant cost savings through better authentication techniques like have been described here.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Just a comment, an anecdote. A couple of years ago, I was in Europe speaking with some senior members of the health ministry in this particular country, and they have a national patient ID. We made some comment about how that would solve some of the problems in introducing a system in that country, and he basically said no it won't because those numbers are traded freely on the street, and they've become worthless as identifiers, which is really the source of my question. Mere possession of the number does not prove that the number actually belongs to you. So that's kind of the backdrop for where I was coming from, but I appreciate your answers.

I will comment to Paul, one of the things that healthcare facilities have in abundance is privacy information about the patient that only the patient should know, so you don't probably even need to go out to public databases to ask questions, that that individual should be able to validate which encounters have you participated in? Which providers have you seen? Which hospitals have you been in, etc.? By definition, we have, in general, a lot of private data that can be used to validate without revealing anything to an attacker or challenger. Thanks.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
Can I make one more just quick comment? Within VA, we do have a unique identifier for all of our patients, but it's not as forward facing as, say, and FSN might be. It is encrypted in the card, the ID card that we do issue. So it can be used for lookup, but it's always verified with other data traits, and the photo is always on that ID, so that should offer some secondary checking.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Yes, those secondary characteristics, I think, that are key. Thanks.

Paul Egerman – Software Entrepreneur
Thanks. John?

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
Very interesting testimony and I have a question that's maybe a little different than what you've spoken to so far. I think one of the problems that I've seen within my organization is that when we have a patient,

and we find inaccurate information, we correct the information. Then our systems that are supposed to be intelligent to match patients often take bad data and try to put it back into the system as being the authoritative information, which then propagates the problem. That sort of leads me to the question, how do we know what is the good data and what is the bad data, especially when it comes to data that's necessary, demographic information that's necessary to do all the matching? What types of strategies are in place to try to make the right decision as to what data is good, and what can be done to try to improve that?

Paul Oates – Cigna – Senior Enterprise Architect
The question I think you're asking is how do you know when good is good.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
And how do you keep good data? Yes.

Paul Oates – Cigna – Senior Enterprise Architect
Again, I think this is an area of best practices that we establish as a community. The approach that we've taken is to identify authoritative sources of information. What that means to us is it's a source of information that we already know and can trace its lineage, so whether it's an eligibility system or something that has source verification. We as an organization know what that system's lineage is and can consider it an authoritative source. That first step is going to be unique.

That answer to that first step is going to be unique for every organization. A hospital information system or registration system will be an authoritative source for a hospital system. The practice management system will be an authoritative source for the physician's office because they trust it. They know it. They're the ones that can control the lineage of the information and where it came from.

The second technique is once we identify those authoritative sources, we also make sure we know that there's identity consistency within that authoritative source. So that if John Smith is ID 123 in that system, John Smith will be ID 123 in that system. If a system does not or a process does not have the capability of being consistent in how someone is identified, we start to question how authoritative it can be because the feed or source of that information needs to have a consistent identifier on it.

And the third technique is anything else other than what you declare as an authoritative source isn't, so it's not used as a contribution back to the relevant data that's considered authoritative. But it is used and linked in as a composite of that individual, so out of 20 records linked to that person's master person index, and I mean the word link, not merge, I mean the word link, 10 of them might be from other sources. They're not considered authoritative, but they're linked there until such time as we get a better answer and they're not. So the other technique is differentiating between authoritative, non-authoritative, and making sure that we're linking and unlinking, not merging.

Scott Schumacher – IBM Software Group – Chief Scientist
I 100% agree with that approach. Merging is probably the most evil thing you can do with structured data. There's probably worse, but I can't think of one right now, so linking of records is key because that traces back to the authoritative source that gave it to you. Then, once you've linked that together within the systems, you need to be able to do a couple things in terms of data stewardship. One is you will make an error, and you will link somebody in that you shouldn't be linking in. So when someone reviews that, and any good system has somebody at the edge looking at data, and they say, yes, that's a false positive. Go fix that.

Once you capture that, the system needs to be able to unlink and keep that unlinked. So the algorithm doesn't try to keep putting it together when a human has said it's not. The same thing when you've put records together, you keep them together. So the system needs to respect the human in the loop. There has to be a human in the loop to do that.

Then also, it's important to keep track. Within our processing is to keep track of historical information, so I get a record from a hospital coming in as Scott Schumacher, and then the next one comes in. It's the

same MRN, say, but now I've got a different address and now perhaps a different phone number. So keeping attribute variation over time is critical for matching because you come in, and you have, if you've been able to put those data together, you can start to link in old data. Linking in old data, so I was kind of scared when someone said they were linking in birth certificates because birth certificates happen typically a long time ago for most of us, and the data on that really is not relevant to—the demographic data really isn't relevant to that.

We did a situation in the U.K. where we were integrating up children's data, and vital statistics was part of it, and that in terms of data quality was the worst. But what you need to do is to keep track of how attributes change over time when you can capture that information, and understand what source is authoritative. So I don't like to say that there's bad data. All data are good. So we keep it all together. We link it all together. We keep all the variables, and then you bring them back out, and you present it from authoritative ….

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
So you're really speaking to a fairly sophisticated record locator service that really needs to adjudicate, manage data and sort of the state and the age of the data or at least the age of the relationship that you're tracking.

Scott Schumacher – IBM Software Group – Chief Scientist
Correct and you need to match across all of those. So if you've got name changes, there are attributes, which change over time. The obvious ones, phone, cell number, etc., like that, address, etc. But also names change over time, and people just decide, I want to use my middle name today, and that's what I'm going to use, and so transpositions and middle name matches to first name matches are common problems that go with this. But as the data change, keep track of it. If you know, if you get an identifier from the source system that says this is John Smith, and you get another one that says here's John Jay Smith, keep those both together and use all of that information within your matching system.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
Thank you.

M
If I could just do one short add on, I think the question about record locator service spawned a thought, and that is, what we're talking about here because I'm thinking this might be a good vocabulary to add to the discussion, so what we're talking about here is a registry style master data management pattern. I think Wes' group and others who have done this kind of analytics in the past can differentiate between a registry style pattern and a transactional pattern. The difference being, a transactional pattern, there's one place and only one place where all data in the universe gets updated, and everybody has to go to it, versus a registry style where everybody gets a chance to update it, but somebody is accountable for trying to wrap it all together.

As we're talking through this discussion, my observation is we're talking about practices, not standards. I think one of the outcomes of this hearing might be ongoing development and listing of things that are good to do in behaviors and practices, but likely not getting to definitive standards. The list seems to be growing.

Paul Egerman – Software Entrepreneur
Thank you. Very helpful.

Richard Elmore – Allscripts – Vice President of Strategic Initiatives
Trust in the authoritative source is really important. We see this repeated a number of times, so it tends, that good registration information tends to flow from the revenue cycle side historically towards the clinical systems. Healthcare organizations, which tend to interact more frequently with their patients, tend to have better demographic data usually if they've got the right systems behind them than those that are more episodically involved with the patients. But then you have an integrated delivery network where the hospital and the physician each think that they are the better source, and there can be some

challenges—you're smiling, John—as you know around data governance. Data governance in a distributed system, it really depends on what the model is, and I think it's beyond the scope of your question, but I think it's a really important topic for the tiger team to consider and address.

I'd also like to just comment briefly on merging is evil, Scott's provocative comment. I think that for linking purposes for health information exchange, that's a good, solid way for the industry to proceed. If you think about the use of merging information into a patient record in an institution where there might be ancillary systems. There might be devices and so on, where there needs to be some patient matching, or there should be if there isn't today, and some ability to integrate that in for the workflow of the physician, there may be other cases where we may want to kind of soften the view of merging.

Paul Egerman – Software Entrepreneur
Thank you.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
I have a couple of additional comments. We also think that merge is evil, but we do use it within our local systems, but at a national level, we only do linking and unlinking. The other comment I wanted to make was about the authoritative source. I think that we've taken an approach where because we have so many medical centers that are very similar, in many ways their information systems are very similar looking that we actually have authoritative sources that will move between different medical centers for a patient during their longitudinal health record for their visits. You may be a snowbird that is seen actively in the south and during the cold weather, and then moves up north, so we actually have different systems being authoritative during different times of the year. Or it could be that we look at different factors such as whether you're an inpatient in a medical center that makes you most authoritative for that patient's information, also if you are providing them with a current prescription, those kinds of things. I think it's important to look at the business flows and what supports that from the business side to what substantiates what is authoritative at different points.

The other point is the metadata. We have business rules around metadata such as if you have a middle initial, and somebody updates it to a full middle name, that we keep the middle name, and it becomes our primary gold record. That you can't change it back to a middle initial unless there's some manual intervention that someone has requested that you don't use their full middle name, which we usually try to deny because we're using it for matching, so those are the points I want to make. Thank you.

Paul Egerman – Software Entrepreneur
Very helpful. Thank you. Alice?

Alice Brown – National Partnership for Women & Families – Director HITP
I wanted to thank you, Paul, for your comments about expanding this conversation not just from patient matching, but to all consumer matching and not just in the care setting, but across the entire healthcare spectrum. I think it's increasingly important and relevant given our increased focus on personal health records and mobile health technologies. And my question is whether you think that if and as we expand this conversation to broader matching across broader settings, and as authentication increasingly occurs, not in a face-to-face manner, the consumer privacy concerns, which we've heard, don't seem to be quite high now, will increase.

Paul Oates – Cigna – Senior Enterprise Architect
I appreciate the sensitivity around the privacy concern, and it is critical because trust is what will allow all of us in the healthcare system to actually change our behavior and get well. We have to trust who we're dealing with. We have to trust who's working with us.

I'm not sure that patient identity and matching is in and of itself a privacy issue. If I look at the generation that I just sent to college, they share lots of stuff that I would never ever imagine sharing. So I'm thinking that privacy can be handled within context of what someone is willing to do. I think what we really need to investigate is to expand the use cases that we're considering from beyond the treatment venue, to expand the lens to the other jobs the consumers want to do.

I appreciated David McCallie's comment that there's encounter data that we could use to verify individuals, but I think they might find that a little unnerving if we shared that with them. On the other hand, there's a lot of things that consumers want to do for which there has not yet been an encounter, like registering for the first time to take a health risk assessment or approaching a physician's Website to log on and start a relationship. Both use cases, which have to be once registered and verified, linked up later again and validated against the repository of information that exists there. I think, if a person, speaking personally, if I think that that person is behind that curtain is acting in my best interest, I'm probably more willing to share than if I didn't think otherwise. Other than that, I think there's a much larger discussion to happen that's probably out of the scope of what I can comment on right now.

Alice Brown – National Partnership for Women & Families – Director HITP
Thank you.

Paul Egerman – Software Entrepreneur
Very good questions, especially when you think forward to the concept of universal coverage, the issue of then the difference between a patient or a consumer and a subscriber perhaps becomes a little bit different. It's a great question. Wes?

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Thanks. Everything is vague to a degree that you don't realize until you've tried to make it precise. I learn that lesson every time we hear testimony. The one thing that I clearly understand is merge is evil. So I think about the following scenario now, which is one that gets tossed around HIE circles a lot.

Various unmerged records are available, and HIE has access to them. The HIE produces a composite of continuity of care documents from that data, sends it to an EMR where it is now accepted in that EMR. Now I like the scenario we heard this morning about a patient came to the clinic with an onset of diabetes based on blood tests and a spontaneous change in race. The physician was obviously being a little droll about talking about bad data.

The change of race would have got lost in that process that I described. In other words, at some point we take this data that has been loosely linked and create an association that wasn't there before. I really don't know what to do about that except to decide how important is that use case to trade it off against evil. I'm interested in your comments, but I have a couple other questions. I'll ask them all at the same time.

We've been talking about accuracy, and I've heard some numbers, 3%, and 8% to 12%. I'm not sure 3% of what is what, or 8% to 12% of what is what, particularly since one of the hardest things I can think of to measure is how accurate is your identity data. As far as I know, the only way to measure how accurate your identity data is, is to go back and manually look at records and try to say that a person is a person … differently than the computer decided on the accuracy.

Finally, to what extent are the numbers different for passive match, which is almost always the scenario in an HIE, versus active match, which is frequently the scenario in a healthcare institution. So the three questions are, what should we do about composite data? What are the accuracy figures you're quoting about? What's the difference between active and passive match in terms of accuracy? Thanks.

Scott Schumacher – IBM Software Group – Chief Scientist
The first is really not a is a virtual merge, which is not as evil as a real merge, but it's evil as well. But you have to think about how you want the system to work. If you're going to do that, you really are doing merging. If you're going to merge, you're probably going to have to run it at a higher false positive rate for that. Other systems that we have seen in fact you do not ….

Paul Egerman – Software Entrepreneur
A lower false positive rate?

Scott Schumacher – IBM Software Group – Chief Scientist
A smaller false positive rate, excuse me. I usually use the negative log of the false positive rate, so I get confused on signs. Fortunately, there are only two. So if you're going to do it in a situation where you are merging, either virtually or physical, then the false positive rate requirements need to be much tighter. I think that, to your comment about there's no difference if you are physically merging it within the database or you're virtually merging it as a composite view.

The other is the way we have seen it done is in fact you don't merge it. You provide out the physician with the information and the information at the record level from where it came. So you don't combine it from information. You say I've got record one in which this guy has some race and diabetes, and I've got record two in which he has a different race and no diabetes. The physician looks at that and says, that looks like an error, and there should be a button on the screen in which he could say that. You can present the detailed information to the physician and have him use that before he puts that into his EHR. But I agree with your comment. If you are going to merge it, and that's what you're done here, then you need to go with a very, very low false positive rate requirement, and recognize that even in that situation you're going to have that.

The second in terms of the numbers, you're correct. The best way to figure out what a matching system is doing is to get two of them that are quite a bit different, give them the same problem, and then spend a whole bunch of time going over the results. Customers typically don't want to pay us a lot of money to do that, and so that's not our process. What we have is we have a process in which we can, basing on profiling of data in terms of the entropy if you want in names and the information you have across that, of estimating what these rates should be as a function of threshold. So we do have a process that we roll through, and we think it's fairly accurate. We've had people that go back and test it at different environments. But essentially it's a process that you take the data that you're going to be matching on. You run it through the statistical process, and you get out of that a curve that says if you think this threshold, you get this false positive rate, and you get this false negative rate.

Then what we have are customers do within that environment is if you decide you're going to live with that threshold, look around it. Measure it. See what's going on, and see if those rates are really what you're achieving or not. So there's a way of estimating it from the data, but finally what you want to do is to refine that estimate, as the system evolves and you get better at it, is to review it with the type of work that Sara had talked about earlier.

Then for passive match versus query match, our philosophy, which I'm not saying is correct, is to use the same approach. We use in fact exactly the same algorithm to match to records as I do if I query. So I just take the query, and I make it look like a new record. So that means that you have to enter enough information to within that passive match, within that query, in order to bring that information. The one difference between whether you're doing a persisted link, which is sort of common in the hospital, or whether you're doing a virtual link, which is on the glass, if you want, putting it together, which SureScripts, as you know does, is that you lose in the latter the ability to do reporting or proactive things. That if you don't, the advantage of maintaining the links is that you can do measurements and reporting upon the data. If you're doing virtual links, then you've lost that ability. That's true really across industries.

When I sell this to insurance companies or I sell it to retails, it's the same issue. If I want to say, if I want to be able to profile my single customers and do analytics on them, I have to maintain links. If I just want to make sure I treat them right when they walk in the door, I can use what we call virtual links.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Scott, I'm going to interrupt for a minute to try to make something more precise here. There are situations where you look at a bunch of information about records, the computer decides that they're about the same person. There are situations where you present information about several sets of records to a person, and a person decides whether they're the same person or not. I call those passive and active matching.

Passive is often used for other than queries, but particularly in the HIE world, it's used for queries as well. There are concerns about privacy that say you don't want to be presenting information about where a patient was seen if it's possibly not the right patient that you're talking to and so forth. I'm trying to distinguish a little bit query from active matching.

M
… filtering.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Yes, attended and unattended, and I'm just trying to get your sense of what are— I guess this is a problem. I have trouble asking the question, but if the circumstances require an unattended match, you use your algorithm to set the accuracy different. How do you deal with that in an unattended versus attended match? I originally was just going to ask how accurate it is, but I realize that the algorithm is the same. It's something else that's different.

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
It's the same. It's the threshold is what you've taken action. In a passive situation, a record comes in. It scores against the records in the database. If it crosses a particular threshold, a likelihood threshold, then you'd link them up, and you're done. If you think about the on the glass integration that I talked about that SureScripts uses, it's the same thing. I go out. I find all the records that are above the threshold. I present them coming back out.

In terms of attended or query type thing, what customers do, many of our customers do is that they adjust the threshold as part of the query. So they say, I would like to see everything—if I'm doing automatic linking at 12, then put in an 8, and let me see down in the list farther. If that's your question, you can query at lower likelihood questions because that's a real time process. Whereas once you've established the link, that's passive.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
The point I just want to make is that that process is much more likely to go on in an institution than it is in an HIE.

Scott Whyte – Catholic Healthcare West – Senior Director, Physician & Ambulatory IT Strategy
Yes. Correct.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Thanks.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
I have a couple of comments on that. How we do that in practice, we do look at attended and unattended searches very differently, and we do use the thresholds. We look at those. But they are different situations. For instance, when we're perhaps matching with National Health Information Network or something, we have a yes or a no. Either you match or you don't. If there's one match, and we know for sure that that is the patient, it's a positive, and we move on. If there are potential matches, it's a no, so we're not saying that we don't have that patient. We're just saying that we can't distinguish that we only have one. That's unattended.

The attended searches, in the past, when we first implemented our VI and we had registration clerks doing lookups on even just a local database, they were presented with a list of potential patients that they were looking. But as you said earlier, they took the top one. Unfortunately, the list wasn't at that point really ordered by the threshold, meaning the closest match was at the top. It was actually the oldest record, which normally wasn't the one that they were looking for. So we've taken that away.

We've taken away the ability for them to choose from a list. I think part of it was they weren't choosing the right one necessarily most of the time. Others were it seemed like it was a bit of a privacy issue where they were seeing other patients that really weren't the match, so we wanted to tighten that up.

Now what happens is if there are potential matches in there, they actually get an identifier, and that goes into a bucket for manual review work. Now the good thing is, now that we have a better probabilistic, more sophisticated algorithm, there are fewer of those, and we make the match the first time more often. But we have really removed that from the clinical setting where we don't have a nurse or a physician choosing from a list of potentials.

Peter DeVault – Epic Systems – Project Manager
Thank you, by the way. Great testimony. One of the things that I was intrigued about was, Scott, you made a distinction between data completeness and data quality. Of course, it occurs to me that different care settings or different points of linkage, those things might be very different, so in the emergency room when a patient comes in, you might have poor scores on both. When somebody is enrolling for insurance, you can enforce that there's at least data completeness, etc.

As the conversation has developed, it occurs to me that the question of what the acceptable error rate is, is not a question with a singular answer. It's one where we need to instead look at different care settings or different points of linkage and characterize them by what's the expected data completeness, what's the expected data accuracy, and what is the acceptable error rate? For that given care setting, what's the required follow-up or review that we need to do afterwards in order to clean that up if it's above the greater rate? I guess what I'm saying is that there's a lot more work to do than might have first appeared, but I'd be interested in your comments on those observations.

Richard Elmore – Allscripts – Vice President of Strategic Initiatives
I think this is responsive both to your question, Peter, and to yours, Wes. I think that there is— We don't know what we don't know about the data inaccuracies. That's for sure. So the use case of drug user who would provide multiple identities, the use case of a family member who provides their social security number for their child. There are many cases where you're going to have errors in the underlying data. Populations that have more uninsured, other conditions will tend to have higher rates of errors, so there are a number of exogenous factors that go into the information, in addition to the points that you raised, which I think are absolutely important ones around the expectations you would have would be different depending on the settings of care.

The measurement is relative to those factors, so it's going to be different in different parts of the country and, within a region, it will be different depending on the socioeconomic areas being served. I think that still there is an ability to, through review of matching that's been done, through human review and use of tools to determine potential for error, to determine whether or not there is room for improvement through error correction, through better processes upfront and so on. I think that really rather than being a— We should be thinking of it in terms of creating a learning system rather than creating a standard that somebody reaches for some incentive purpose. I think that if we do that well as a healthcare community nationally, we're going to have a very large improvement in terms of patient identity and patient matching.

Paul Oates – Cigna – Senior Enterprise Architect
Peter, I think you're spot on, and I really like Rich's characterization that this is not a topic that we can set a standard for. This is a topic that we create a set of best practices and a learning system for, so we elevate the awareness and the skill set across the industry in dealing with these topics because it's a skill based and best practices based and process based discussion. So if we could train and raise the awareness of this topic to the same level as many of the other kinds of topics we have in the healthcare exchange discussion, I think that would be very valuable. So I think your observations are absolutely accurate.

To also return briefly back to specifically to Wes' questions about the lost data when we're doing composite CCDs, one of the practices, again, not a standard that we took in our organization is that when multiple parties are combining data, they're not allowed to do that in and of themselves. They first have to reach out to that place where that entity exists and codify their information with the current best composite answer for that person and tag their records with that. In the HIE model, that would presumably require in a process before you put information together, reach out to the HIE and find out

who you're doing it for. So there might be some use case development there to figure out how best to do that.

You also asked about the privacy piece. The process we've put in place when information across multiple domains gets linked and some of that information is sensitive is that we have a post process that takes that result and pings it up against the policy. It's in a system that says who can see what for what purpose. If the requestor is not supposed to see that information, it's redacted appropriately, so that it's a two-step process. We don't mix the privacy piece and the linking piece. It's a two-step process, as a post process.

Then lastly, what are we measuring? What's the numerator and denominator? We're measuring overlap between sources that source A and source B share this many percent of the same people that look like duplicates. We're measuring fallout. We don't know if it's a false negative or a false positive, so we're going to have somebody look at it. To be frank, that's a bit of a rig to percentage because, based on resources, you're going to control your false negatives and false positives. Once we set that in place, on an ongoing basis, what we really measure is data quality and completeness. In stream, as data flows, we've put processes in place and tools and systems to watch what goes through the pipe. If some things over time trend and starts to degrade, maybe a new person got hired or a system put in a release, and suddenly the date of birth is blank now or some identifier that was consistent is now scrambled, we get a warning, and we have a closed loop back to that process to find out what happened.

Scott Schumacher – IBM Software Group – Chief Scientist
I really do agree with Rich's comment on this is a learning system. I think that perhaps I was a little bold in the beginning by saying we ought to establish rates, but I think what we need to do is start to measure what we have and how it's going. I agree with your comments that the systems have to be different for different care settings, and they also have to be different for different uses of the data. But I do still think that there's a key issue between, and maybe it's my sensitivity, between data completeness and data quality because data completeness, as a vendor, I can't do anything about. Data quality, if I'm missed a match because my algorithm isn't very good, my customers kind of yell about that. So that's why I developed French phonetic algorithms and other components to it is to go around these data quality problems. Your algorithm grows to take care of that, but data completeness is again sort of a fundamental stop.

Paul Egerman – Software Entrepreneur
Peter, are you all set? Great. Joy?

Joy Pritts – ONC – Chief Privacy Officer
We've heard a lot about the need to measure and how this should be a learning system. We also heard in the earlier panel that there are lots of even large organizations that currently do not have this capability. So I'd like to know what types of organizations. Should there be any limits on who should measure, or do you see this as something that is the responsibility of every type of provider?

Scott Schumacher – IBM Software Group – Chief Scientist
Yikes. I think that might be slightly out of my wheelhouse. What we ask our customers, anybody who is using the system, is to measure it. We do not have measurements of people who are not using a system, although they're there, right? As Scott Whyte had said in hospitals that aren't, that are sort of normal, the normal error rates you're going to see in that is 8% to 10%. But the only way to get that is to run some sort of a matching algorithm on it.

Who should be responsible for doing the measurements? As close as you can put that to the owner of the data, the better the system will be. So if you were saying to hospital A, you're going to participate within this HIE, then you measure your data, and you show me that you've got the data quality that's going to make it—a data quality and data completeness that's going to support the RHIO. I actually think that you have to push it back to the owners of the data to do the measurement or to be corrected to the measurements. I don't think you can correct it at the center of an HIE.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
I was just going to agree with Scott. I think that it's part of the responsibility that you hold as an organization or as a group if you enter into these sharing agreements or relationships with other organizations that if you are going to be linking records, sharing data, that you owe your patients or your customer base the fact that you have done some research. That you hold yourself to some kind of a standard for your error rates, and that you establish those based on industry standards, other areas of the industry that have done it before you. I just think that it's the responsibility of the data owners to do that, and I imagine that not everyone would, but I think that that would be a recommendation.

Joy Pritts – ONC – Chief Privacy Officer
What about the small docs?

Paul Oates – Cigna – Senior Enterprise Architect
I guess I was going to offer that the aspect of linking data across stakeholders means that someone is sitting in the middle looking at all that information come together, and that someone, whether it's the HIE or whomever, probably is in a position to measure and enforce an accountability model. But I do think that along with adoption of electronic health records or any system, including dating back to when we all bought practice management systems, comes with it the obligation to get the data right. So even the small doc doesn't have to spend a lot of money on sophisticated tools, but probably does need to train their staff that if you have a field and it requires a value, you put it in. Hopefully the systems that are being purchased and certified have some foundational level of data completeness that doesn't let a transaction go in or allow someone to get around it without putting those criteria in place, and so that's part of probably the training and development process. I'm a little out of my wheelhouse too here, but we do it in our organization too. If you are an owner of a process, you are also the data steward of that process.

Richard Elmore – Allscripts – Vice President of Strategic Initiatives
Joy, I appreciate the question. I think it's a really important one. The scenario where we have a significant gap in healthcare in the U.S. today that the smaller providers have not, up to now, been making the investments in the technology that would be supportive of more sophisticated matching techniques. A lot of them are using deterministic matching where they will have a set number of fields, which have to match in an exact manner. It's a little bit of an oversimplification, but it is a problem out there in the industry, and it's something that I think we do need to address going forward to make sure that all of healthcare providers, all the healthcare stakeholders have access to reasonably advanced matching technology to insure privacy and safety or respect.

Paul Egerman – Software Entrepreneur
We have a few more questions, but we're also running short on time, so what I'd like to ask for the questioners is to do their best to be brief, and I'd like to ask you also to be brief and point out it's not necessary that all four of you respond to every question. It might be a way we could also get in as many more questions as we can. Next will be Gayle.

Gayle Harrell – Florida – State Representative
I'll be very brief. Joy's question was so pertinent. I thank you for it. It brings out really who sets the accountability standards and where are they going. Who has got that responsibility? Where does that buck stop? What is the role of transparency in this? The public has a right to understand and know what's behind that curtain, who is running the show behind the curtain. So I really would like some comments, please, on who sets—if we're not going to go to standards, if we're going to measure and be accountable, who is going to set and make that apparent to the public, and where does the buck stop?

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
We'll keep it really short and nobody will respond. I think it's a really great question, to be honest. I think, as a person who consumes healthcare, like everyone is, I'd like to know how my healthcare is matching my record with others. I don't think it's something that we, within VA, have really done yet.

I think that we feel we are accountable for the matching that we do within VA for our veteran patients and for the other patients that we treat. But I don't know that we widely publicize that outside of our technical industry, and I think that maybe we would look for you as a recommendation on how we should be doing that in a consistent manner, and what is the mechanism that we do that. How much do we need to let the public know at a level of information that they'll understand what it really means? It might also help with some of the privacy issues and why am I giving this information, and why is it that I had to give you all of this information last week, and now I can give you just a few traits? Well, because we've matched your record and we don't need to take the full spectrum. I think that it's a really good point.

Paul Egerman – Software Entrepreneur
Deven.

Deven McGraw – Center for Democracy & Technology – Director
So we've done a little bit of work on identity on the Internet of individuals and finding ways. We've put forward recommendations for what we call user-centric identity systems where at least in that Internet space where the individual is the one sharing data in accordance with his or her preferences they might choose which identity points that they want to expose in some sort of knowledge based query system or otherwise. But they would have that choice versus systems that I've heard talked about today where essentially you ping the ... that you think are authoritative and people don't necessarily have a choice in what those are. Maybe they're public most of the time and so this isn't really an issue, but Rich also mentioned the sort of voluntary choice of adopting an identifier and certainly, I think a lot of us have read about it.

So there are a couple of pieces to my question, but I'll make it quick. One is what is the potential out there in healthcare for some user-centric identity where the patient offers information and creates sort of an identity algorithm that then can be sort of persistently used? Then second, what if there is such a voluntary system, like the voluntary identifier, but you have a patient who chooses not to use it? What are the backstops to making sure that person still can be appropriately identified if they have privacy concerns about choosing a number?

Paul Oates – Cigna – Senior Enterprise Architect
Very briefly, the aspect of a consumer creating their identity if that's, I think, what you were saying, I put some comments in my testimony asking the same question. We have seen the variability around consumers creating their e-mail addresses and it's a disaster because there's no consistency and you can't identify who you're sending e-mail to, so taking that to the broader healthcare system I have some concerns about all of us independently creating our identities and how to have consistency around that. As far as the solution, I don't know, but the pattern seems to be established in the past about some concerns around that.

The second around value of a voluntary identifier, the approach that I've seen is that if you have more information we'll be able to use it. If it's not there we won't use it. I think one of the challenges of a voluntary identifier is not just the creation of it, but how to propagate it and get it operationally implemented across all of the various stakeholders, because there are many systems out there that may not even have room for yet another identifier to store it and use. So the national provider identifier I think might be a use case we can study about how well that's gone and see if there are some learnings that we could re-use for the consumer side.

M
I think that's a great question. I think it sort of gets to the point of we've really been talking about a B2B system here, right? Hospital A to hospital B, physician A to physician B. When you talk about B2C, you're really in a very different world.

We have multiple retail customers, who try to address that problem where they have an authenticated identity for a particular individual and then maybe they have multiple other authenticated identities and they link that together. It gets relatively complicated in keeping track of what's going on within that realm, but the basic model is this sort of registry system where when you're coming in through an authenticated

way you really only see that narrow point of the domain. Then when a user is trying to come in, the company is trying to come in from the other end, they tend to see the entire person. So you've linked these authenticated identities and they can be multiple. I could own six products from somebody and be linked together, but the company treats it different from the consumer, but I think that that's a real complicated, sticky area.

Rich Elmore – Allscripts - Vice President of Strategic Initiatives
Deven, the question about a user centric identity systems: There might be some lessons that can be learned from folks' experiences with personal health records and patient engagement strategies. Where providers have had to make decisions about how they look at information, clinical information that is provided by the patient, which they typically tend to view as important sources of information, but not validated in the same way than information that might have flowed from within their own organization. So there might be some lessons to be learned in that area that can inform us on your question.

Paul Egerman – Software Entrepreneur
David.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
I have a very specific question for Sara. To the degree that you're able to answer it, I was surprised a little bit by your relating that you no longer show a list of close matches to the clerks, who register a patient or a patient coming into your system. It seems to me that that introduces a lot of risks that you might have a mismatch and not recognize it. I'm curious. What do you use to drive the direct match that has to match before you go forward? Is it based on the ID number?

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
It's based on a set of ID traits that are collected for patient lookup; the full name, date of birth, gender, Social Security number if we have it.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
And if that's not a perfect match then a human—

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
It doesn't need to be a perfect match. We use probabilistic matching, which weighs on those traits and nicknames are matched. I mean Scott could probably talk for hours about how the probabilistic matching works, but it's because of the success of the matching algorithm that we're able to do that.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
The probabilistic will always match more than one person. The other persons will just be at lower probability—

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
Right, we set thresholds.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
But you only show the top one?

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
Actually, I misspoke. We show the ones that pass the threshold that we've set, so it may be a couple and they're in the order of the threshold—

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Okay. That makes much more sense.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
Yes.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
I just didn't follow it.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
But it's a high threshold, because we don't—

David McCallie – Cerner Corporation – Vice President of Medical Informatics
It's a high threshold. Right.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
Right.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Okay. That makes sense. Thank you.

Paul Egerman – Software Entrepreneur
Thank you. Actually, that was very helpful understanding that. We're out of time, although I still have one last question. In our very first panel, the question was asked if there was any benefit to standardize or normalized name formats. The people in the first panel said yes, there would be some incremental benefit doing that. Do you agree?

Scott Schumacher – IBM Software Group – Chief Scientist
No, I don't.

Paul Egerman – Software Entrepreneur
You don't think so?

Scott Schumacher – IBM Software Group – Chief Scientist
No, I don't think so. I think that there is a lot of variation in the way people record names and if you think about a name standardization process that's a binary issue—I mean a unary issue. If you think about matching, that's a binary issue. So if I get a name that says it's Al Smith. Do I know it's Alvin, it's Albert, the rest of that? When I'm comparing two and it's an Al against an Albert that's when I do it. So, standardizing names before you match is, I believe, an incorrect thing to do. Standardizing addresses before you match I think it is, because usually there's a knowledge source involved.

Paul Egerman – Software Entrepreneur
So standardizing addresses around like the postal services—

Scott Schumacher – IBM Software Group – Chief Scientist
Correct.

Paul Egerman – Software Entrepreneur
You think that would be beneficial?

Scott Schumacher – IBM Software Group – Chief Scientist
Yes. It is beneficial. It's measurably beneficial.

Paul Egerman – Software Entrepreneur
True.

M
Paul, could I ask a clarifying question on that? Standardizing of names I agree with you. What about name fields, so, for example, middle name, middle initial and suffix? I run into the problem. I'm a junior and every time I go through TSA screening if my driver's license says junior and the airline reservation doesn't and it raises questions and I can't auto-validate my ticket, because nobody handles junior consistently—

Scott Schumacher – IBM Software Group – Chief Scientist
Right. So in terms of blocking of names, we ignore that completely. So if you put your first name in the middle name or you can put it in the last name. You can put junior up front. We go through and we figure out what we think is in the name, so people hyphenate it or they'll use the first one as part of the middle name. That's all part of the algorithm. That goes to the Mathematician ... Employment Act I mentioned earlier; is working through those variations. But it's really a comparison of two things, so if I've got three name tokens from this record and two name tokens on this one, I'm going to do all of the comparisons and take the best sort of alignment out of that.

M
... is a better tool.

Scott Schumacher – IBM Software Group – Chief Scientist
Yes. Right.

M
It's pretty clear that the airlines that I fly don't use your software.

Scott Schumacher – IBM Software Group – Chief Scientist
They don't. They don't. We've had that discussion with them.

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
I just wanted to ask Scott, so don't you think though that if it was possible to put something out there as a standard for what would be the way that you would like to get the information? To say you should create or accept a suffix, prefix, full middle name, blah, blah, blah? Because I think within organizations we should strive to do that and I just wondered if—

Scott Schumacher – IBM Software Group – Chief Scientist
Well, I think my standard would say put down everything you know, right? So if you have a suffix I want to see that. If you have a title, we take care of that too. You don't make much weight out of it. The most important thing you can do is make sure you get a middle name instead of a middle initial. I mean in terms of elevating likelihood that's a huge thing, as you know, specifically in the U.S., but that's critical if you're talking about matching on cultures where the last name space isn't as rich as the United States. So getting away from first name, initial and last name and going to as many names as you've got, throw them all in, it improves the process.

M
Does the fact that you asked that question, Sara, indicate that you think there would be some benefit to some standardization on makes or perhaps to the extent that David suggested?

Sara Temlitz – Veterans Health Administration – Data Quality Business Product Manager
Yes. It's in our written testimony. I think within our organization we've really strived to do that, to use the ones that exist nationally and those that don't, to develop them for VA for use. We use those if they're not standards, formal standards, we do have policies that we put in place and directives that come from within VA so that our registration clerks are all operating off of and clinicians, anyone entering names operates off the same policy and standard to get the most complete information we can.

Scott Schumacher – IBM Software Group – Chief Scientist
If I were to do any kind of integration with Sara, what was most useful for me is to know that she did that so I know what to expect versus another organization. The fact that she did it or not is immaterial to me, but to know that she did and how to consider that as part of the usefulness would be ....

Paul Egerman – Software Entrepreneur

Thank you. Very helpful. Terrific presentations. I thank all four of you for your presentation. We are now going to take a break until 1:00 and then we'll resume at 1:00 with panel number three and continue to discuss about the solutions.

I think, Deven, did—?

Deven McGraw – Center for Democracy & Technology – Director
Yes. Only to say for those of you who are not familiar with the area, there are actually lots of lunch options. There is, in addition to what's available in the hotel, if you wanted restaurant style there is a COSi right out the door. There's a Top T-Salad one block down. Lawson's Gourmet. I did my thing here—

Paul Egerman – Software Entrepreneur
Boy, this is impressive.

Deven McGraw – Center for Democracy & Technology – Director
... on 13th Street in the same block. Lawson's Gourmet, also on 13th Street. So don't go far.

(Overlapping voices.)

Deven McGraw – Center for Democracy & Technology – Director
Don't have any French Indian Fusion. Sorry about that.

(Lunch Break)

Judy Sparrow – Office of the National Coordinator – Executive Director
I think we're ready to begin, if everybody could please take their seats.

Paul Egerman – Software Entrepreneur
Good afternoon and welcome back. We are going to continue with the panels on what are the solutions and Deven and I are going to flip around. She's going to manage this panel and I will be the timekeeper. I would just comment that so far this morning there was an extremely impressive track record in terms of every speaker nailing the five-minute limit on presentation. It is unbelievable. In fact, Secretary Sebelius was aware of this and the challenge is can we do that this afternoon. Can everybody do that this afternoon? There is an additional burden on you. We're all looking to you to absolutely nail that five-minute target, but I will be watching to make sure that we do that. I'll let you continue, Deven.

Deven McGraw – Center for Democracy & Technology – Director
I think for those of you who weren't here in the very beginning, Paul set the bar very high and we enforced it pretty vigorously. There were some penalties associated with failure to do so that were, I think, clearly understood to be less than desirable, so—

Paul Egerman – Software Entrepreneur
You'll be on the no-fly list.

Deven McGraw – Center for Democracy & Technology – Director
Yes. One of them had to do with tax audit or something.

Paul Egerman – Software Entrepreneur
In the afternoon, it's the no-fly list. If you miss the five minutes—

Deven McGraw – Center for Democracy & Technology – Director
Somehow.

Paul Egerman – Software Entrepreneur

We're going to put your name or something that looks like your name on the no-fly list and so you will be here in D.C. for another three or four weeks.

M
But feel free to talk as long as you want.

Paul Egerman – Software Entrepreneur
That's right.

Deven McGraw – Center for Democracy & Technology – Director
On that note, I'll say that we have clearly a very distinguished panel here. There is one name on the agenda of someone who had a last minute conflict and was not able to join us and that's Paul Contino from Mount Sinai and the Smart Card Alliance, so more space for you all at the table. We have Mark Gingrich and Paul Uhrig from SureScripts; Ken Tarkoff from RelayHealth; and Sean Nolan from Microsoft. We'll go in the order of the agenda so, SureScripts, you're up and you get five minutes total.

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
Perfect. Thank you. Well, Mark and I are going to tag-team you today because Mark won't let me talk about anything technical, so I'll start off. You all know who and what SureScripts is, so clearly, the topic of our discussion today is our ability to identify and match patients to their insurance eligibility, benefit, formulary records and to be able to provide med history to a requesting healthcare provider. All of the requests are pursuant to the NCPDP standards as adopted by CMS.

At a high level, prescribers submit to us patient identifying information: name, date of birth, zip code and gender in a medication history or eligibility request. Social Security numbers, license numbers and similar identifiers are not used by SureScripts to identify patients. We use an MPI then in order to match that information to the records maintained by PBMs and pharmacies. You've heard a lot, no doubt, about MPI today, so I won't belabor that, but our MPI is a collection of membership and patient demographics that's provided by each payer, PBM, health plan and pharmacy that participates in our network.

When a request for med history or eligibility is made, it returns either a patient found or a patient not found response. Our MPI is based on an algorithm that's tuned to our specific needs and accounts for certain things for, by example, in a mistyped field. So, for instance, the algorithm we use has the ability to conduct derivative searches so that if a patient's name does not precisely match that in the payer or the pharmacy, searches and comparisons are done on variations of the name.

If the comparison score exceeds the match threshold that we have set the software produces a match result using assigned weight values that produce the high probabilistic score indicating that the return records are indeed that of the patient being requested.

If a patient is found then in the case of an eligibility request the MPI matches the patient with a payer assigned, unique ID number for payer information and forwards the eligibility benefit request to the PBM. The PBM uses the patient demographics and the payer assigned, unique ID number to confirm the coverage exists and to retrieve and return to the prescriber the benefit coverage information. In the case of a med history request, if a patient is found the MPI matches the patient with database keys that are used to retrieve the med history record from the participating pharmacy. All such records are then transmitted back to the requesting, treating provider. If the patient is not found then a message is sent back to the requesting provider stating that the patient was not found and no data is available.

Our MPI is structured to match patients with a high degree of certainty, so we analyze approximately 50 million PBM patient benefit records to determine the fields and data elements that most accurately match patients. We've also conducted extensive research and maintenance of the appropriate ... vaccine or bucketing of patient information for this functionality. We also regularly test the MPI for quality, conducting test cases for accuracy and testing the algorithm.

I think you had a testimony earlier about dynamic linking versus static linking. We have structured our MPI to dynamically link patients to their records. In addition, under our program, once a patient is identified our application certification requirements require the display of demographics and benefit information for confirmation by the application user, so if in fact there is a difference, maybe a name spelling between the request and the response that is highlighted to the prescriber.

We actively seek to reduce any potential false positives and false negatives by seeking updated information from payers. We may receive updates on up to 10 million to 20 million members a day, especially in high enrollment periods. In our view and experience, the algorithms that we use, our false positive response is extremely low, so we've had about ten false positives out of approximately one billion requests in the past nine years. So we believe that uniform standards for name translators and updating of source information would improve matching processes and getting updated data on a timely basis. Thank you.

Deven McGraw – Center for Democracy & Technology – Director
You'll end up with some time during the question period to put some more stuff in and, of course, we have your written statement as well. Ken.

Ken Tarkoff – RelayHealth – VP & General Manager
Good afternoon. My name is Ken Tarkoff. I'm the General Manager of Clinical Solutions for RelayHealth. RelayHealth is a division of McKesson Corporation. I want to thank you for the opportunity to testify today. At a high level, RelayHealth Clinical Solutions is a business that offers clinical connectivity technology to health systems, health plans, provider organizations and consumers. Additionally, we're one of the largest players in the regional and community health information exchange space and one of the few that has a patient as an active, on-line participant in that exchange.

I think it's clear to everyone today from the testimony that there are inherent challenges with the data when addressing patient identity. There are known issues with data quality, data completeness, standards and formats. We also know that there is inconsistent use of identifiers across health systems, provider organizations and in connecting an exchange. We also know there is often manual intervention required by either locally or centrally in a central organization or out in a practice in order to resolve those matches.

Also, it wasn't talked about much, but we know that patients generally do not feel accountability for participating in helping to resolve this issue. One of the things that I don't think was mentioned today is we believe there are significant legal and financial risks to providers for inadvertent disclosures of mismatched data, particularly as stage two and stage three requirements start to roll out into the market.

There clearly are many approaches that were discussed today in the market to address patient matching. We tend to define them in two different ways; one around centralized matching, which is when you can use a centralized identifier or process for resolving the identity. We also have decentralized matching, which is when you're using an exchange; you're actually doing the matching out at the locations of the exchange when you're not required to have a central identifier in between. At RelayHealth, we use both approaches. In a decentralized world, we're focused on leveraging clinical workflow for the provider to give them incentive to actually resolve the exchange, for example, using referrals or orders or results management as a way in which in order to address that exchange.

From our perspective, the recommendations we have for resolving this problem is, number one, to ensure a minimum standard data set for matching. We believe that if you required standardized, demographic data fields and formats to use we will actually reduce the amount of errors required in the matching and also increase the likelihood that when manual intervention is involved that the success rate is easier to accomplish.

We also believe, based on the premise that humans will always be involved and errors will be created and if patients don't feel the accountability to monitor, measure and fix the data then we believe that we need to put a flag or set flags inside the data in order to exchange when you identify the data is bad.

Recognizing that there will be errors, we need a process to ensure that that data doesn't get promulgated throughout the system and that there are ways in order to resolve that data.

Third and most important: We believe it's important to balance the workflow requirements on the provider to ensure a high enough bar of confidence on the match and then also to make sure that the workflow fits within their environment. Today there is significant financial risk that will be put upon the provider in many markets if there are errors made and that data is exposed from a PHI perspective. We believe it's very important to balance that in the marketplace and it's something we're going to need to consider in the market going forward.

So, in summary, at a high level we believe that this patient identity matching is a solvable problem. We believe that you can start from a good enough place by getting standards around data matching, standards around formats, an ability to identify data when there are errors and ensure that there is appropriate workflow to resolve. We also want to make sure that that balance for that physician is recognized so that we don't get a resistance from the physicians in adopting the technology as we pursue this further. I appreciate the ability to participate in the dialogue today and I'm willing to answer questions later on. Thank you.

Deven McGraw – Center for Democracy & Technology – Director
Terrific. Okay. Sean.

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
Hello there. Thank you for the opportunity to have the discussion. I'm Microsoft and I certainly degree it's of significant importance. We deliver products across the care continuum and every one of them is dependent on this patient linking challenge at some level, including our enterprise products, Amalga Versions, and our consumer health platform, HealthVault.

While there's a great deal of discussion that can be had around techniques for computational matching, it's pretty clear and I think we've heard this today, that there's really never going to be a perfect solution to the problem technically. That even ignoring their political implications, unique identifiers also don't provide a complete answer, because ID cards get lost and people make mistakes and we still need to deliver care. So we believe what's most important in the national dialogue is to recognize that different use cases require different levels of confidence and that it's important to build this recognition into the standards and requirements that we talk about so that we don't accidentally suppress innovation and positive impacts that could otherwise occur. I'm going to highlight two examples that we see in the real world and the different ways that we approach the problem.

A key tenant of the Amalga system is that we try to use the right information at the right time when questions get asked and it demands we retain a great deal of information about the information that we manage. For example, the degree of confidence we have in matches that have been made and how those matches were made. With this metadata available, we actually can do the right thing in different situations. For example, when matching biometrics, like an EKG, a really high degree of confidence is required and if unique identifiers don't produce something perfect it's very appropriate and probably required that humans should review that before any clinical action is taken. But when presenting a list of patient allergies, also a clinically very important problem, it may very well be appropriate to accept a higher false positive rate in order to reduce the chances of false negatives in that case. Because obviously, the ramifications of missing a potential allergy tend to be much more severe than treating around one that may not actually exist.

In these cases, you might also use UI. We talked about that earlier today to actually show the doctor these differential confidence so that if there is additional investigation that should be done that's available and apparent. Obviously, when you look at aggregated information and population metrics, very often error rates can be statically washed out and so lower confidence rates or experimental techniques may be appropriate in those cases as well.

In all of these what's most important is not really how the match is computed in the whole, but how the metadata around the match is used to support the appropriate use case, and each one really is different. We think that doing this and keeping that data also has secondary effects. It allows us to enable improvement over time as things advance. It supports off-line validation better. It helps us generate training sets for other algorithms and it's very effective in encouraging the kind of resiliency and recovery and error fix rates that Rich and others have talked about today.

I'll turn to consumer health platforms, like Microsoft HealthVault. They introduce another level of complexity into this problem, which is already pretty good and complex. For example, demographics that you've got through the Internet are generally self-asserted and difficult to trust as input to a matching algorithm. At the same time, in those cases anything less than near perfect matches is unacceptable because of the privacy and legal implications releasing sensitive information.

Finally, there are significant privacy aspects here of actually sharing unique identifiers across loosely coupled boundaries, particularly around collusion. If one service knows that patient X has HIV and another service knows that patient X lives at a certain address those two people have some very damaging information together.

So to address the issues, HealthVault has actually rejected traditional patient matching for our own internal match system. Instead, we rely on something we call dual credential presentation. The idea there is that if an individual can prove that they own a particular HealthVault record and at the same time that they are a specific patient at an institution, the link between those folks can be inferred and HealthVault doesn't need to be involved in fuzzy matches as it were. Now, that doesn't completely solve the problem, because each connecting institution still has to decide how to best identify their patients, but they can do so in a locally appropriate way, from very conservative, in-person techniques to more on-line friendly, private report style stuff, as we've talked about.

We also deal with the privacy risk very explicitly and intentionally by giving each external system its own randomly generated identifier to reference an individual record. So while two hospitals may both have relationships with the same individual in HealthVault, each knows that patient by a different pseudo-identifier. In that way, HealthVault is able to act as a de facto EMPI and encourage exchange between these loosely coupled systems that really have no knowledge of each other, but we don't introduce this same risk of collusion. I appreciate the opportunity to chat very much and look forward to the questions and going forward.

Deven McGraw – Center for Democracy & Technology – Director
Okay. Great, many thanks to all of you and thank you also for the testimony that you submitted in writing. It's really helpful for us to understand the complexity of this issue way more than you can possibly do in a five-minute span, so very much appreciated.

Now, as I said, you get more opportunity to get your points in through the questions. I'm going to start to the right, since I started to the left last time. I'm sorry, John. So therefore, I won't be accused of having some right hemisphere defect thing. David, go ahead.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
It would be your left hemisphere affects the right visual field.

Deven McGraw – Center for Democracy & Technology – Director
Right. Correct.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Remember your neuroanatomy. So I came in just 30 seconds late and, Paul and Mark, my question is for you because you may have already covered it and I just didn't hear you. In SureScripts and RxHub's merger, you have taken systems that had different databases of patient data. When you were referring to the examples you gave of the 50 million members and so forth, were you describing a merged version of those two or the old RxHub PBM data? Are there differences in the approaches basically?

Mark Gingrich – SureScripts – SVP & Chief Information Officer
Right. That was the initial analysis that we did back with RxHub when we talked about the 50 million member records that we used to define the algorithm. Since then we have, through again ... testing process we have actually migrated to one algorithm that we used across both.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Across both?

Mark Gingrich – SureScripts – SVP & Chief Information Officer
Yes.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Then my follow-up question would be given that you have kind of your definition with the four data elements that you mentioned of kind of the gold standard, are you able to disclose how many false positives there would be if you had perfect presentation of that four data elements? I mean just a simple select statement from your database, select, unique, etc. How many false positives would there be? In other words, how many patients are not differentiated by the use of the four identifiers? You don't have to answer this, but just if you have that data.

Mark Gingrich – SureScripts – SVP & Chief Information Officer
Yes. Again, the statistics are used to define and refine the algorithm, take us to the level that we've seen today. I think we talked about the probability of a false positive is right around what we're seeing, which is one in a billion or ten in a billion. But no we haven't; I don't think we've done any analysis to say; again, we have a database that has the member records that we use for the match. We have history of requests that were made, except we don't maintain the PHI long-term for those requests, so I don't think we have done that analysis that you're talking about.

Deven McGraw – Center for Democracy & Technology – Director
Gayle?

Gayle Harrell – Florida – Former State Legislator
My question really is, and I want to make a comment first, is to Ken. I'm so pleased that you've brought up the liability issue, which, of course, I had brought up earlier this morning. I see, as we move forward into the world of HIE and exchange, extensive exchange, which is what we all envisioned coming out. I see that liability issue growing and becoming more significant, so I think that's a policy issue that this group is going to have to discuss at length. Thank you for bringing that up.

My question really goes back to the issue of minimum standards. You brought up the need for some minimum standards. Do you see ONC putting out a rule on minimum standards through a certification process for whom? Would you certify HIEs? Would you certify vendors? How would you do that and what are some of those standards that you would want to see incorporated?

Ken Tarkoff – RelayHealth – VP & General Manager
Thank you for asking that question. One of the biggest problems in health exchange is you're relying on the data that's sent from the sending system and the quality of the data that's being sent. When you're going outside of the health system, outside of a closed environment, from one to another that don't use the same centralized identifying logic and approach you're forced to use demographic data.

Unfortunately, what happens today is there are no standards even for basic demographic information. Some systems don't provide zip codes when they send the information. Some systems don't actually used generalized standards. I believe for any amount of exchange that happens you need to have the sending system, the systems that are required to hook up both to send and receive need to be able to follow those minimum standards.

So the specific answer, I would say, is yes. Anybody that's participating in that exchange, any system that's required to participate that's going to be certified in order to do that exchange should, at a minimum, set a certain number of demographic fields that are agreed upon. A standard for sending that data along with fields that are identified for when problems of data are identified, when there is a mistake or an error or an issue that's identified by a patient or by a provider, a way to ensure that that gets notified. Because when you send that data into the HIE and it gets promulgated throughout the system and then you identify that there is an issue, that data may be spread throughout a variety of places and you need to have standard sets for the way in which you expend that data.

I don't know if I answered your question specifically, but I think it's very important any system that has to certify that they send and receive data has to send those minimum data standards from our perspective. Which demographic fields?

Gayle Harrell – Florida – Former State Legislator
Yes.

Ken Tarkoff – RelayHealth – VP & General Manager
Well, that's actually a good discussion, because there are a minimum set of five or six fields we use, like name, date of birth, zip code, address and gender. Those fields are very important from a minimum set of being able to send them.

Probably one of the things that would be good to look at is see how many of those you can get that's standardized. For example, one of the things I didn't have a chance to talk about is we do see issues with sending systems with special characters and what actually gets created with middle name. So you might want to look at middle name ... in how you address middle name and initial, because sometimes they merge it into the first name and when they send it on the system the way the system sends it is they can break it up. Now, you can resolve those with probabilistic matching, but then the liability goes up depending on what bar you set. So that's why I was saying it's important as you figure out the field. You're not going to solve it; you're just going to reduce the amount of workflow that's required to resolve it when you're sending it across.

Gayle Harrell – Florida – Former State Legislator
A follow-up question: How would you deal with the HIEs that are taking that information out? Would you set that taking the information in and perhaps then transferring it if they're doing composite, if they're linking or the negative on merge as opposed to linking or merging or doing a composite picture of all of the information they have in? Would you set certain thresholds or standards or certifications for HIEs dealing with the information?

Ken Tarkoff – RelayHealth – VP & General Manager
Yes. I definitely think that HIEs need to be able to be certified to be able to accept that data as much as anyone who is involved in the exchange. They need to have a minimum set of standards and how you would certify that they can support those standards because there are a lot of systems that touch this process. So you're talking about all of the systems along the way, so how you'd actually execute that from a certification process I probably can't comment on, but there are a lot of systems that it's not just the provider organization. You also have all of the points of contact, the ancillary care providers, reference labs, the patient side as well too; all of those components that are extending the data back and forth need to have a standard set and if you could do that through certification I think that would be a good mechanism.

Deven McGraw – Center for Democracy & Technology – Director
Thank you. John.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
I have two questions, one for Paul specifically, since you're really relying heavily on payer assigned ID numbers it sounds like. Assuming that we end up with a national healthcare environment where everybody is covered by somebody, do you believe we could stitch together what amounts to a unique

identifier based upon the fact that in theory they're assigned to a provider who would have a unique identifier with them?

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
Actually, just to make sure it's correct; we have a unique identifier for the member. That identifier is assigned from the source, so that's just identifying that information and that record at the source. So as far as coming up with an overall unique identifier—

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
I'm just trying to think of a way to stitch together what would be a unique identifier for everybody based upon something you know from an individual payer; I'm sorry—

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
An individual payer?

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
Yes.

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
Again, you have to remember that there could be co-coverage, so even within our eligibility—

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
Right.

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
There are multiple payers that may provide information back in a response.

M
....

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
Right.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
That's true. Yes.

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
Right. Exactly.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
I had one other question. I have a question for Ken. You said that patients do not feel accountable for ensuring that their information is accurate or something. I'm paraphrasing you. I don't know if you were here for the earlier testimony just talking about the totality of this environment. What do you think the patient role should be and how do you think you try to accommodate that knowing that you have a very broad environment with a lot of different types of players in it, who need to try to match patient information?

Ken Tarkoff – RelayHealth – VP & General Manager
Excellent question. I'd actually take that in two parts. One is I think one way I would answer that is there has to be a fair amount of education to the patient about—the consumer. Excuse me. It really is the consumer. It's not the patient, education of the consumer of their role, that they play a role in this and that they need to be active in revealing the data and ensuring it. I can think of the equivalent of how many times you get advertisement on TV to go in and check your credit to ensure that all of the information in there is correct so it doesn't negatively impact your credit—

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security

Right.

Ken Tarkoff – RelayHealth – VP & General Manager
And how you could actually get consumers to feel obligated to go in and check their data.

The second part of that, and I think it's in line with what I am anticipating we're doing in stage two and three is requiring providers to provide that transparency to the consumer so they actually come on and see their data. Those two things lined up will help that.

One other thing, as I mentioned as well too, is we really need to be careful about the potential liability associated with these metrics. You need to educate the consumer to go look, but for example, in the state of California the Medical Information Act is set where if there is a disclosure, regardless of harm, the patient is obligated to $1,000 per incident. So if you're in a situation as a provider, who is doing this patient matching and we're using all of the technology that's been talked about and all of the services in doing everything that's standard and doing a really good job we all agree it's not going to be perfect. There is manual intervention, which will have errors. Each time there is an error that could disclose it there is a financial risk to that provider at that point in the state of California. That could be a real problem to achieving our overall policy goals as an organization. We have to think about how that will impact what we do as we deploy the technology.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
Do you believe it shifts the burden then? I mean I'm asking. I guess it does shift the burden to the patient to some degree, but I think there are going to be people that will, regardless of how much you try to engage them, are simply not going to.

Ken Tarkoff – RelayHealth – VP & General Manager
I agree. I don't ever believe there's one answer to this. I believe you've got to hit it from all sides and you have to continually make progress. What I worry about is significant barriers that would stop adoption of overall technology. So a lot of physicians are nervous about putting their data on-line because, A, they trust theirs more and they're worried about the patient seeing what they have, quite frankly, because they know that their data is not all 100% accurate, so there is that concern.

If you change the paradigm, the direction that we're going, and we get more patients to get involved and we get people to pay attention now that there's that transparency I think you'll see the shift in behavior from the providers recognizing that the consumers expect to see their information and they want to see it. We have to make sure that they don't feel an overwhelming burden of liability as they start to get involved with this, because that will then make them hesitant to adopt any technology.

M
Could I answer that just really quickly?

M
Sure.

M
There tends to be this sort of sense that patients really don't want to be part of this or they don't go out of their way. Our experience has been that's actually not quite true. People are very compelled and interested and desirous of being involved in that care and being part of those exchanges. It's just really difficult today. The additional burden for that mom, who's taking care of kids, is very high and that creates an inability often of people to get engaged. As we look forward and as providers are obligated to share with patients and as repositories or ways for people to manage that becomes more automated, I expect you'll see patients playing that role very naturally and enthusiastically when it's not being asked of them in terms of we're just going to dump a bunch of problem on you, but we're now giving you a set of data that will allow you to be a better parent or a better patient.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security

I could see in the case where I want to see what my record looks like, I want to see what's in my record. What I thought I heard Ken really say was something different, which was I have responsibility for ensuring that the data exchange and my identification across providers and potentially insurers is accurate, which is really a different thing.

M
I actually see them as an extension of the same thing though, because effectively what people— When we have discussions with people and we talk about the fact that you really understand the gaps in care that might exist because of assumptions you have about who is making the decisions with full data versus not and the reality underneath that. By sharing that record and being able to say here is the data from these various places and seeing those mistakes, people actually can take a lot of very positive action without any extra work. It still is an extension of seeing that record. It's seeing it in a combination though that there is a lot of differentiation and mistakes there.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
So they really only see the linkages once the linkage has occurred. I guess what I mean by that it is I guess a patient couldn't query the system to see how they would match if somebody were to ask for information.

M
That's for sure. That's fair.

Ken Tarkoff – RelayHealth – VP & General Manager
John, if I go back I'm glad you made that point because the way we define a patient's record is not just one provider's view of their chart to the patient. The patient actually navigates through the system, sometimes in and outside of closed health systems or they have dependents as well too, so their record really is a much larger thing than one particular view. So seeing all of the places of where they have data that's being provided and all of those points of linkage that have been created is important for them to see. Because now they can actually participate and say, —Hy, I never went to that physician," or, —That's not my information," and notify them to resolve that issue before it actually becomes a problem. I actually think that will reduce the errors as well too, because obviously they know their own care as well.

Deven McGraw – Center for Democracy & Technology – Director
Wes.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
I've got a question for Mark or Paul and then for Sean. So you described that you have a large database of patients. I didn't hear the actual number because I came in late, but I'm going to guess that it's getting close to 100 million patients. Is that—?

Mark Gingrich – SureScripts – SVP & Chief Information Officer
Two-hundred forty million.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Two-hundred forty million. Okay. So I was safe in my guess. A little low though. You only use four data elements to identify them—

Mark Gingrich – SureScripts – SVP & Chief Information Officer
Name. Date of birth. Name field, so including prefix, suffix—

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
Yes.

Mark Gingrich – SureScripts – SVP & Chief Information Officer
Name. Date of birth. .... Zip code.

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
And zip code.

Mark Gingrich – SureScripts – SVP & Chief Information Officer
And gender.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
And gender. Okay.

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
Although gender doesn't have a lot of information after name, does it?

Mark Gingrich – SureScripts – SVP & Chief Information Officer
Right. Exactly.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
So you then talked about measures of accuracy, but one of the questions I had was when you say I don't have a match; that is when you make a negative determination how do you know that that's not a false negative or do you know? I mean we—

Mark Gingrich – SureScripts – SVP & Chief Information Officer
We don't know, actually. In the analysis that we did up front, again, when we put the algorithm together with Mr. Schumacher here –

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Sure.

Mark Gingrich – SureScripts – SVP & Chief Information Officer
We erred on the side of the false negative, right?

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
And I forget the original analysis, but it was a significant potential for not returning information—

Mark Gingrich – SureScripts – SVP & Chief Information Officer
No. No. I think, obviously that's the right answer—

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
For e-prescribing.

Mark Gingrich – SureScripts – SVP & Chief Information Officer
Right. For our purpose.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
The question though I have is there's sort of an implied or actual assertion that we only need those four data elements and I'm just wondering have you any sense that with more than four data elements you would be able to—

M
More precisely—

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Be more precise in—

M
....

<u>Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst</u>
Yes and so what is your answer?

<u>M</u>
That's the part that I was trying to ask as well, albeit, not as eloquently as Wes did.

<u>Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst</u>
Right. We have not taken that analysis on. Actually, based on the up front analysis from the peers that the top three were part of the original analysis, those are the only data fields with a high enough predictability and high enough quality for us to actually put the algorithm together at that time. We couldn't rely on the security field, again, because in some cases it was not available based on system regulations within the information that we were allowed to have in eligibility. So those were the data elements we could work with. I would assume we could get to perfection if we had additional quality data elements.

<u>M</u>
But we're not talking about perception, perfection here. You can dismiss any argument by saying that you're trying to be a purest, right?

<u>M</u>
But we are talking about most of the other people that have testified have said, ―You give us data. We do the best we can. You give us more data we can probably do different." You seem to be denying this and—

<u>M</u>
No. I think, again—

<u>Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst</u>
... the standardization. I mean we're talking; we're having conversations here about creating standards for data that enforce new data collection requirements on providers. If there's only this many data elements then it's not such a big deal; if there's more then it's more. So it's a pretty critical issue. Obviously, there are probably no other data elements that are uniformly available. Social Security number doesn't have a check digit. Social Security numbers are used by multiple people and there's all of these digits.

<u>M</u>
Right.

<u>Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst</u>
The only question is is it somehow additive? Is it worth being in the equation or not? What I hear is we don't know.

<u>M</u>
I know that there are papers out there, studies that—

<u>Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst</u>
Your initial data was across PBMs, right?

<u>M</u>
Yes. Right.

<u>Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst</u>
So they're kind of remote from the patient anyways. They don't ....

<u>M</u>

Right. So I would say, again, if there are additional fields that would have quality and predictability we definitely could improve that negative side.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
It seems like a likely hypothesis, but the truth is we don't know, right?

M
Right.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Sean, I wanted to drill down on your comment about; and I'm paraphrasing; we get an actual match. We don't use fuzzy logic to attempt to determine a match. I think probably most people understand, but just check me here: I have a HealthVault record. I have a record with Walgreens. I have logged into Walgreens and HealthVault at the same time and at that point you know that the person who identified themselves by WRishel in your database is, in fact, the person who is Wesley Rishel in Walgreens. Dead solid fact, because I was ... to both systems. I was authenticated in both systems when it happened.

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
Exactly right.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Right. You don't even know that I'm really Wesley Rishel. All you know is that I knew how to ....

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
You have those two credential sets at the same time and you're alleging or making the allegation that they are the same individual.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
And if I go on and do that with Qwest or with Mayo or someone else then you accumulate a set of languages to me, which you carefully don't give a common identifier out for. Okay?

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
That's exactly right. It's a hub and a spoke model.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Right. Yes. In a lot of ways that sounds like the ideal method. There's no talk about false positives. You don't talk about false negatives. You just say it's the patient's responsibility to do this.

The question I have is what do you do about the last mile, which, to me the analogy of the telecom last mile program is all of the patients registered in all of the different doctors' office systems that are out there as opposed to all of the ones that use Walgreens, the big providers.

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
Absolutely. Well, this actually is an interesting balance for us because there is such heterogeneity out there that we provide many different ways to sort of help providers create that situation such that the two credentials can be put together at the same time. The tension point is that there are many and it can be complicated or hard to talk about at times.

One though that stands out very typically, a classic way to do this—I was just speaking about this with some others—was that at the time of an office visit when an identity is well believed to be as solid as it can be. There is a mechanism where a system, whether it's a fully connected EHR or just through the Web can actually acquire a token from HealthVault. That token can be given to the patient physically and later used in a process where they log into HealthVault or create accounts at home and then present that token to simulate in that off-line sense the same kind of exchange. That's the typical case for somebody who comes in for treatment and, as much as I would want it to be the case, nobody has HealthVault accounts today ... numerical sense. Ask me again in a year, but—

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
I asked last year, so—

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
We're not there yet, but in that case the workflow is critical that we're actually able to start the experience there and carry it through to the data and the event that actually just happened. That's one of the mechanisms that we use. But there are many, many and it's, frankly, one of the biggest challenges. We believe patients being part of that as the norm of registering for their service and having, whether it be HealthVault or GoogleHealth or any of our systems. That dynamic has to become naturally part of the provider exchange and then we really do have that hub and spoke ability to be perfect with our exchanges over time that I think avoids much of this challenge, at least for loosely coupled systems.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
A couple of points: One is that you've made it much easier to establish that linkage from a provider's office than it used to be with this token. Two, I think the reason you were saying ask me in a year was because of meaningful use incentives that are coming into play and direct—

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
There are some great initiatives happening this year that I think will make things look different next year. Yes.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
Okay, I mean because I think the fundamental question is motivating the provider to do that—

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
Absolutely.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
And if there's a way to do that then we can look for this approach of direct being more fruitful.

M
.... We actually used the same methodology that Sean described when registering patients on our service. We allow it's called an invitation that's sent by the provider's office with a token for credentialing. Often where you're capturing an e-mail address of the patient with the information you have some series of questions and you can get them on-line. That's a common, very good technique for bringing people on-line where you have a certain level of authentication of who they are. It's an important way, when we look to later on in stage two and three, how systems are going to need to use that type of technology to bring consumers on-line to exchange with their providers, which would then lead them to be more active in their healthcare.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
So, in both cases when you say a token it could be as simple as a string of characters that they type into the computer? It's not a physical—

M
....

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
That's correct. It's all done on-line. A series of questions that you ask the consumer based on information you've collected that they can identify themselves.

Deven McGraw – Center for Democracy & Technology – Director
Okay. Paul.

Paul Egerman – Software Entrepreneur

First, I want to thank everybody again. This is a fascinating discussion, especially, Wes, that last discussion was very interesting. I actually just have a very specific question for you, Ken. I read your written testimony and there's one part I didn't quite understand. You had some recommendations about implementing processes for correcting data and for standardizing transaction sets or flags around incorrect data. I just wondered if you could explain that.

Ken Tarkoff – RelayHealth – VP & General Manager
What I described earlier when I was talking about one of the situations where in a manual flow in a provider's office where they get a set of probabilistic match and they make a decision that is errored. I actually described in my testimony a situation; the first wife and second wife; that can happen. Those types of situations do happen, particularly when you have manual intervention. There's enough data that makes it close enough or potentially the process isn't exactly locked down. When that match is made and that information gets shared, particularly as we become more and more networked, that information then can be flowing through the system without the ability for people now along the chain, maybe even from a trusted source because situations like that happen in very trusted environments. Now you don't have a way to undo what happened.

One of the things that we need to do is set up a process for identifying when there is potential for flagging. It doesn't necessarily mean because if the data is linked it doesn't mean that all of the data in it is bad, but it does mean that along the way the amount of trust that you put in the data needs to be recognized in the system as quickly as possible. So you could do it through ADT feeds. You can do it. There are a variety of different mechanisms in which you could identify fields to capture that. I think we should look at the ways to make sure that that gets identified so we don't have a system when we know data happens through errors that we have no way of notifying throughout the chain.

I'll give you a specific example. In that situation where you mismatched it and then prior to identifying it, meaning let's say the wife actually called and said, ―Hey, you've got the wrong person," and told the provider. By the time that wife got on and looked at that information that practice had already shared a health summary out to the HIE that had gone to one or two other places. At that point, there should be a resolution process in the technology to be able to identify there is an error to be able to send the message through the HIE and know where the data has been sent to be able to, as quickly as possible, identify that there is an error. Then when it gets resolved, to then be able to change that flag the same way, and so there are a number of ways to do that today. There are some profiles that exist and there are some things that we could do, but I think it's very important we look at that because three or four years from now if we don't have it in place I think it will be a problem.

Paul Egerman – Software Entrepreneur
... you're looking for like more than one thing, but one is like a transaction set that notifies certain people downstream or upstream or something that an error occurred, that a mismatch occurred?

Ken Tarkoff – RelayHealth – VP & General Manager
... field inside of an existing transaction, so you can use a field inside of an existing transaction to be able to identify that field.

Paul Egerman – Software Entrepreneur
I'm curious. Are you trying to notify people that the mismatch has occurred or are you trying to notify them of what data is incorrect?

Ken Tarkoff – RelayHealth – VP & General Manager
Well, ideally you would notify them the data is incorrect and correct it, but sometimes just because someone says, ―Hey, there's an issue," it may require research and let's say the research takes a couple of days in order to resolve the particular issue. I think there is an obligation every day, every minute that goes by that someone else is using the data. It's potentially being used in a different source. We need to be able to flag that quickly to make sure people know that there are potential issues is important.

Paul Egerman – Software Entrepreneur

But to do that kind of a process, flag people with incorrect data, don't you really have to define who is the authoritative source? Isn't there a risk that you're going to create a situation where everybody is notifying everybody else of data being incorrect and—

Ken Tarkoff – RelayHealth – VP & General Manager
Is there an authoritative source for data today? I guess I would answer that question by saying when you're in a fully networked environment and you're connecting all in a community or in a region, lots of different health systems, lots of different major players. A patient that goes to a health system, a large health system if you figure they have the best infrastructure for it; if they identify a patient and there is an error that is made they're a trusted source in the exchange. But there was an error that was made and they provided a health summary to another health system because that patient went to a physician splitter for example and that happened. So now you have a situation where it is a trusted source, but there is no one source of data that can say to everybody this is, in fact, true. So that's where the complication is. I'm not aware of; I've seen in these types of environments that you could say there's one source that ultimately knows all.

Paul Egerman – Software Entrepreneur
I agree. It depends on the environment. I mean the information exchange environment you described is like the hardest environment. There are some environments where there might be a trusted source—

Ken Tarkoff – RelayHealth – VP & General Manager
Absolutely.

Paul Egerman – Software Entrepreneur
I think about the panel we had previously where we had a presentation from somebody from CIGNA. I look at it as if CIGNA says this is the correct subscriber ID number, well that's a trusted source, because they know their subscriber ID numbers, right? So if they want it corrected the provider ought to go with whatever CIGNA tells them to do for that particular field. Now, whether or not the provider will necessarily go with whatever that person or that insurance company tells them for an address I don't know whether or not they would consider that a trusted source, but they might not.

Ken Tarkoff – RelayHealth – VP & General Manager
Yes, but the thing I would just encourage you to think about is we all talk about it's not perfect. Even though it's a trusted source, mistakes are made and if people rely on that trusted source and are using that data in other places there aren't mechanisms today to be able to ensure that people are aware there is an issue, particularly when it comes from a trusted source. So if you don't put any infrastructure in place to address that you can create a bigger problem.

Paul Egerman – Software Entrepreneur
You don't think the current transaction set has an infrastructure to do that though?

Ken Tarkoff – RelayHealth – VP & General Manager
I think you can leverage fields with it. The question is, just like in any data field, there is lots of fields, but actually getting some consistency of the fields that are used and expectation that you can take advantage of those fields, I think that's important. I mean our space now with using CCD as a standard and being able to go a particular direction is making exchange easier, because you don't have debates over which type of data field to use and what type of HL-7 structure you use. So those types of things actually make exchange better, because you're creating decisions and standards. This is another example of what data fields to expect people to use in order to do the exchange.

Paul Egerman – Software Entrepreneur
This is helpful. It's really an interesting issue, because there's almost none of this going on right now where one side tells the other about corrections and so this is an interesting issue. I still don't 100% understand how to do this right, but it is an interesting issue.

It also seems to me, if I'm hearing this panel right, there is some consensus that standardization around formats or demographic fields would be beneficial, that that might be something that this Tiger Team might recommend. Is there anything else we should do besides this feature we just talked about, corrections? Is there anything else that you think we should be doing?

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
One thing is maybe best practices around just updating the data for data integrity. I mean as we mentioned, we get millions of updates a day and so I think updating the information to make sure that it's consistent across the different organizations—

Paul Egerman – Software Entrepreneur
When you're talking about updating is that the same issue that Ken was talking about, corrections? When you refer to updates is that what Ken is talking about?

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
I think Ken is talking more clinical care errors. I'm talking more about enrollment data, someone changing their name or other areas that may be the in record that are corrected. That's what I'm talking about, so that you have the appropriate match.

Paul Egerman – Software Entrepreneur
Okay.

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
I would suggest a liability challenge is also something the government could really actually work on. We've had similar discussions around research and anonymization where anonymization techniques are imperfect by their very nature. Matching is imperfect by its very nature and if we want to have the benefits of taking advantage of that technology and having it be widespread there has to, in some sense, be relief. We can't also on the other side put draconian punishments on the people who were doing their best to implement them and do a good job. This is a challenge that I think only government at some level can really address and just say if a set of standards are being followed, $1,000 per incident for doing the best the industry or any human knows how to do. I think at some point there has to be a choice whether you want the exchange to happen or not and this committee is in a great position, I think, to at least have that discussion happen.

Paul Egerman – Software Entrepreneur
It's an interesting issue, Sean. I'm sort of smiling as you talk about the research side, because when we talk about information exchange it seems like we're struggling with this problem that we can't figure out who the patient is. But when we talk about research and we try to make it anonymous we're struggling with the problem that we can't hide who the patient is.

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
That's right.

Paul Egerman – Software Entrepreneur
It doesn't seem fair.

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
Well, Paul, you opened up the can of worms even further if you think about if you provide that transparency and I know there's been a lot of discussion from the Tiger Team about a patient choosing to opt out of exchange. There is an assumption in that that exchange is an entity that is separate, that is connecting systems. Whereas, if you're using the Internet today much the way we do, it's not a different system that's actually doing the exchange. So as you provide the transparency and the liability issues we're talking about the reason why we feel it's important to address that is if there's a huge movement for patients to try and remove themselves from the exchange you're going to remove yourself from the care delivery process, which gets very complicated, because those lines blur pretty substantially.

Paul Egerman – Software Entrepreneur
I mean that's true, but still, there are a lot of people who believe in individual autonomy; if that's what somebody wants to do they should be able to do it. But when you talk about transparency could you elaborate on what you said?

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
So as a patient; and that's what we were talking about earlier when John was asking his question; for example, on our system the patient can come on-line and see all of the points of connection and who's matched them, all of the providers they're connected to, whether they're in that same health system or not. You potentially see that there's a match that was, as we talked about in the identity matching, made that isn't accurate. If you look at that data now you can actually not just see your medical information, but you can actually see who is looking at your data and where you're data is connected, where else you're getting data from. If you lose confidence in that and you decide to say, —ey, there is an error there. I want to remove myself," that's a non-trivial thing to the ability for many physicians or providers to deliver care, because they become so interconnected with their systems. It's not necessarily a distinct entity that is moving data back and forth, because there are lots of different technology approaches for it.

Paul Egerman – Software Entrepreneur
Those are useful comments. I was thinking about transparency in the context of this discussion. Do you think it's important to have transparency about the algorithms that are used for matching? Should providers and vendors be encouraged or possibly required to explain just how does this all work?

M
That's a good question. Go ahead.

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
I would say a the very least being prepared to have that be part of the metadata of the exchange is very useful, because then the receiving party at least has the information they need to take action based on their belief. If they see, for example, a geeky example, generally when you send encrypted data you send the decipher by which it was encrypted and the bit link, not only so that you can decrypt it, but so that the recipient can decide if they trust that as something that was effective at doing that. So if I know that, for example, algorithm ... on the Internet has been known to have a super high error rate I may not accept data that was matched that way.

The question of whether or not they should require or dictate is a very difficult one for the government, but the one that says be transparent about that information and allow exchanging partners to make decisions is maybe a little bit easier.

Paul Egerman – Software Entrepreneur
That's an interesting response, but I look at this hearing and say Microsoft has given us a lot of information about how you operate and SureScripts has done the same thing. So does that mean that ... an expectation of transparency that organizations should explain how they do patient linking and patient matching? It's not an unreasonable thing, at least to suggest as a best practice.

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
Strictly from a transparency point of view it does not seem onerous at all. I was reacting more to the idea that you might try to dictate from the government what would be appropriate with that particular use case.

Paul Egerman – Software Entrepreneur
Do you have any reaction to that, Paul?

Paul Uhrig – SureScripts – Chief Privacy Officer, EVP Corporate Development
No. I mean certainly one of our tenants is transparency. So I think to the extent you can increase trust in the system by transparency of the algorithms, I mean they need to be understandable for the person who is receiving the information, obviously, but I think that's a basic tenant of what we need to try to do to increase trust in the system.

Paul Egerman – Software Entrepreneur
It does strike me that the information that you've given is extremely useful. I mean it's sort of like when people look at what you've done and say, ―Well, that's a best practice. This is something we can model ourselves after." So that's very helpful.

Deven McGraw – Center for Democracy & Technology – Director
Are you good?

Paul Egerman – Software Entrepreneur
Yes, I'm good.

Deven McGraw – Center for Democracy & Technology – Director
I don't put my own card up because I know who I am, but the only other person with a card up is someone who's had a chance to ask a question already, so if you don't mind, David, I'm going to squeeze mine in and then we will get right to you.

So I have a couple of sort of clean-up questions and I just want to make sure I'm clear in my mind. Paul's question about transparency on algorithms and we sort of got a response from all of you. We don't have the other panelists up, so I might just put this thought out there to get your reaction and then if some of the other folks wanted to respond off-line.

We've raised this issue with respect to de-identification methodologies and the issue of being transparent. But it also occurs to me that in some context your particular secret sauce for effectively de-identifying or algorithmically matching may in fact be a trade secret that you don't necessarily want to share all of the details with to the public because that's the core of your business model. So am I missing something or is that potentially an issue? Because one other way around it is to be open to the extent that all of you have been, but not necessarily require people to drill down to the Nth degree in an open, public forum versus if there were an audit of practices that was done without exposing trade secrets. So trade secret; am I just off base or is there something to be explored there? What would be ways to increase transparency and accountability for doing this right, without necessarily risking creating barriers to innovation and incentives to innovate in this space?

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
I'd go back to the cryptography field again and when I think about this it's very similar work. Typically, transparency around those algorithms is essential for the trust in them, for people coming to rely on a tax against those services as a way we learn and improve them. Yet, there still is the ability to patent that technology. That seems a reasonable way to approach that idea, something where the public good is important and yet, some way to the research costs involved, auctioned in those investments are key as well. Obviously, that's a big part of Microsoft's business is our patent portfolio and why we feel that we're comfortable making those kinds of disclosures if they're important.

Deven McGraw – Center for Democracy & Technology – Director
Okay. Ken.

Ken Tarkoff – RelayHealth – VP & General Manager
I guess I would answer I think that's a good point. You're right, Deven, because I actually think there is a level at which you go on transparency before you would get the trade secrets. I think if you think about standard data fields that you used, The data fields are required. The process that you go through on matching, providing that level of transparency. You probably won't get much sensitivity, but actually, what's behind the algorithm that actually comes up with the calculation is probably something that would make a lot of organizations uncomfortable about trying to share something that's a core component of their business.

Deven McGraw – Center for Democracy & Technology – Director
Did you guys want to answer that too, one of you?

M
I'm just checking with legal counsel first, right?

Deven McGraw – Center for Democracy & Technology – Director
Of course.

M
I'm not ....

M
All right. Again, probably more of the meta-level; I think we're free to talk about the level of false positives versus false negatives—

Deven McGraw – Center for Democracy & Technology – Director
Right.

M
So I think we can put some ... around it, but not probably down to the specifics of the algorithm because in many cases we probably can't because we're actually a customer of ... or IBM.

M
....

Deven McGraw – Center for Democracy & Technology – Director
A voice from the peanut gallery.

M
But the process of having something does require disclosure.

Deven McGraw – Center for Democracy & Technology – Director
Yes. Right. That's right. So then, the second question I want to clear up is the issue of the role of patients in helping to correct records. I just want to make clear, I want to make sure that what you're talking about is the transparency of data that we're sort of pushing through meaningful use leading to a more open environment for patients to correct versus some obligation on the part of patients to actively monitor and correct their data. Maybe that is what some of you were talking about, but I want to make sure because, certainly, in the credit report context we are encouraged to check our credit report, but we're not required to do so and failure to do so is, of course, the risk that you bear as an individual consumer down the road. But I just want to be clear on what you're asking, what you're recommending.

M
Yes. The way you described it is correct. We're not implying that you would require consumers. We just want them to feel some accountability because of the impact that it would cause, not require them to do that. I think the credit report is a good analogy, which is clearly, if you go to buy your card and you haven't cleaned it up that's going to hit you. Maybe it's more serious in the clinical world because of the risk, but any more accountability that they feel today I think would be positive.

M
I would agree with that.

Deven McGraw – Center for Democracy & Technology – Director
Well, so I think we're trying to create the conditions under which patients can feel more free. That it's easier for them to see their data, whether that then leads to some greater sense of accountability I just want to make sure that when you're saying the word accountability you're not asking us to hold patients accountable in some way for reviewing your data.

M
No.

Deven McGraw – Center for Democracy & Technology – Director
Okay. All right. David.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
The fact that you were able to remember who you are, Deven, means that you don't have amnesia either, so you're doing really good.

Deven McGraw – Center for Democracy & Technology – Director
Thank God.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
You're doing good.

Deven McGraw – Center for Democracy & Technology – Director
I didn't have to pay for any of this either.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
FreeCreditReport.com, right? Free consult. So my question is a trade secret as well. I want to address it to Sean and Ken in particular, but of course, any of you could respond. Those who have been forced to listen to me through these Tiger Team meetings through the last year and a half and Standards Committee meetings before that know that I'm an advocate of the consumer controlled health record as the proper point of aggregation of the lifetime record. For reasons, which I won't go into here, it makes a ton of sense.

Sean, you correctly pointed out that one of the barriers of doing that is that it's just hard to do. So you proposed Microsoft's approach for claiming an encounter into your PHR, your company's PHR. Ken, you described a similar approach for claiming an encounter into your PHR. Cerner has yet another similar approach to claiming an encounter into our PHR. I assume Peter could describe one. Rich could describe one for our respective companies.

I wonder is there an opportunity here for development of a standard around claiming of accounts, such that providers would feel less barrier to have to do complex relationships to lots of PHR vendors. Consumers would have well understood mechanisms for essentially stepping through the process that the two of you described. I'm encouraged by the experience that a number of us here in the room have had with the Direct Connect Project, where competitors got together and voluntarily hashed out a secure messaging standard, which we're in the process now of validating to see if we've got it right or not, but it so far is an encouraging process. I wonder if there's a similar process that might make sense to do around this notion of connecting provider offices to PHRs. Sean, do you want to take a shot at that?

Sean Nolan – Microsoft Health Solutions Group – Chief Architect
Sure. We get to be somewhat altruistic about this this year and for the next many years, I think, because we're trying to build a market around this. I have not encountered anybody in this space who sort of feels that they're going to protect their market share today.

I actually look at Direct as the primary means that we have at our disposal right now to start doing that. In fact, I've seen VA and many people be excited about it because they've gotten to a point of wanting to share with patients and one way of doing it. Now, there are, as we were discussing earlier, circumstances; it only handles some circumstances now. Anybody who can get a direct address can use the same mechanism and that's exciting. How I go get that direct access is maybe the next step of that game. I think that you'll see that continue to play. I believe probably at this point trying to foresee how that might play out before Direct gets a chance to have its first steps is a little challenging, but there is absolutely momentum and motivation on the part of the vendors that have been participating to try to make that happen.

M
David, I'm not sure I fully understood Sean's answer. I don't know if I fully understood what you were asking on the PHR side. Are you talking about standalone PHRs or are you talking about just the ability to exchange? Because from a real health perspective the PHR is really the patient's view and their information throughout the system where they have access. We actually leverage connectivity with HealthVault, much the way that Sean described. We support that as well too, so I'm not sure if you're asking if we would support—can you clarify? I'm sorry.

David McCallie – Cerner Corporation – Vice President of Medical Informatics
Yes. Great. I apologize for not being more clear. I was referring to ... independent patient controlled records, the health bank style model or the HealthVault style model where the consumer gets a copy of the record from the provider regardless of which EMR the provider has. So it's not the tethered record, consumer view into the provider's EHR, which is an important subset of capability for consumers, but it's a different one than I was referring to. So the notion that if I establish an account with an independent PHR vendor that there's an easy way and a standard way. So that regardless of which provider I go see and which EMR system they happen to be running, I can register and make it easy for that provider to upload the results of that encounter into my PHR without having him to worry about, ┼don't know how to talk to that company or I don't know how to talk to this other company."

M
Right. So from our perspective we're fully supportive of that. I mean we try and do that today, like I described. We do that with HealthVault and others. Absolutely. Great answer.

Paul Egerman – Software Entrepreneur
Great. Mark or Paul, feel free to comment. I think, obviously, you have access to a very important source of patient data that should be made available to those consumer health records, so you probably would need a similar kind of model even though you're a different kind of provider than the EMR provider.

M
Right. Right and we have relationships with Microsoft and others in order to facilitate that.

Deven McGraw – Center for Democracy & Technology – Director
Okay. Dixie, did you have any? Okay. We're a little ahead of time. I'm sort of a little bit surprised. So we'll take a five minute break. Thank you, this panel number three, very much for your very insightful comments. We very much appreciate it. Please feel free to stay for Panel Four if you'd like; otherwise, stay warm.

Paul Egerman – Software Entrepreneur
Are we back? Okay. Welcome back. This is Panel number 4 and Paul Egerman and this is a discussion about approaches in other industries. We have two great panelists here. One is Laurence Castelli, who is the Privacy Officer from the Customs and Border Protection Department of Homeland Security. I just want to say all of those comments I made about the no-fly list I was just kidding. Just kidding. And Timothy Boomershine from Fair Isaac.

We will be consistent with what we've done in the past. You have five minutes each for your opening comments and then we'll have a discussion.

Deven McGraw – Center for Democracy & Technology – Director
Well, just to say we're really serious about five minutes, just so you know. There will be plenty of time during the question period for you to get more in, but we've been really strict with the other panels, so even though there's only two of you, in fairness, we're going to hold you to the same.

Paul Egerman – Software Entrepreneur
Thank you, Deven. So, we'll be given to you, Mr. Castelli.

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
Thank you very much and I'm okay with the five-minute limit as it is. Good afternoon and thank you for the opportunity to address your panel on this topic. To help translate it for what I do maybe what I should do is just quickly tell you my name is Laurence Castelli. I am the Privacy Officer for Customs and Border Protection. We're in the Department of Homeland Security. Our responsibilities, among other things, are to essentially secure the border. Basically, any person coming across the border, essentially attempting to be admitted into the United States must pass through customs. What we do is we seek to identify them and determine their nationality and citizenship, as well as their identity before we admit them into the United States. This is true of U.S. citizens, legal permanent residents, as well as foreign national travelers. What I've done is try to confine my remarks to that interaction of the mission, simply because I think that might offer some insight into how what we do relates to linking of patient records with identity.

I am the CBP Privacy Officer. I've filled this role now for five years. I was the first privacy officer for customs and border protection. Among my responsibilities are to ensure privacy compliance. I prepare privacy impact assessments; system records notices. I am involved in all sorts of fun meetings regarding various system development in new practices at the borders, new practices. Any time we touch information invariably it involves collecting personally identifiable information from persons, either foreign nationals or U.S. citizens. The Department of Homeland Security, by policy, does not administratively distinguish between foreign nationals and citizens. We accord them the same rights, so we're always looking to protect the information we collect from persons and ensure that it is used for the purpose that we collect it and share it with people who are using it for a similar purpose and not allow it to be accessed by others, who don't share that purpose.

CBP employs over 58,000 people. On an annual basis, we are responsible for ensuring the security of over $2 trillion worth of import transactions. On a daily basis, we interact with 1.1 million people crossing our borders. Each of those persons we collect information from. That's, in a nutshell, what CBP does.

Looking at the questions that you had identified, just real quickly, when you cross the border, again, going back to that border interaction, we ask that you provide us proof of your identity and proof of your citizenship. We normally look for you to provide a travel document of some sort, a passport, obviously, is the most obvious form of this. But even providing such a document and I'm sure that other panelists have raised the issue. The mere fact that you have a government issued document that attests that you are who you claim to be does not mean that it is, in fact, a valid document. Obviously, we encounter a fair amount of fraud. I'm not going to say it's a significant amount, but there is fraud with regard to passport and with other types of travel documents. We do encounter, through our officers and agents in the border patrol, we encounter persons who don't cross at legitimate or defined border crossings, who are seeking to evade presenting identity. Again, we are still are charged with identifying these individuals. We look for whatever form of photo identification they may have and whatever form of information might be available.

In addition, we maintain many databases where we have information, both in terms of what you might colloquially refer to as lookouts; information about persons who engaged in criminal activities, so wants and warrants, or information about persons, who—

Deven McGraw – Center for Democracy & Technology – Director
You've got like three seconds.

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
So at any rate, we use these databases to link to various other agencies as well to help handle that data exchange.

Deven McGraw – Center for Democracy & Technology – Director
Yes. I think we're likely to ask you more about that linking thing during the question period, okay?

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
That's fine.

Paul Egerman – Software Entrepreneur
Tim.

Timothy Boomershine – Fair Isaac/FICO (Finance)
Great. I'd also like to thank the committee for the opportunity to discuss some of the issues and solutions to linking our records. FICO is a leading provider of predictive analytic solutions for financial services. Basically, we provide decision support models for most of the major banks around account originations, account management, fraud detection, revenue cycle management. We deal with a lot of privacy concerns. We deal with a lot of the legal concerns related to that and basically we deal; most people don't know it; with almost every aspect of their banking relationships.

To do this and to build these analytic tools we essentially are using multiple large, diverse data sources, both internal data assets that FICO has built over many, many years from different sources, as well as linking out to external data sources. Through this process, we have had to deal with linking records quite extensively and have really addressed the issues in a number of different ways. Although the primary factor that drives investment and improvement in that process in our business is financial for the banks, the issues are almost the same and the factors, although not financial, are probably just as pertinent within healthcare.

The areas that we really focus on and look at is essentially; and a lot of the areas have been mentioned today; data capture, data quality, data completeness, data edit. One of the things that was discussed earlier was whether it was important to have standardized names or not important to have standardized names. Both of, I think, the opinions that were presented were correct. One is not important in terms of probabilistic matching. The algorithms, the tools out there are very good at dealing with that. In terms of a culture of establishing accurate data collection and ensuring that people care down to the level of the person taking the data it is important. Where we don't have that culture, we don't have accurate data we get garbage in. Sometimes we get bio-fuel out; a lot of times, we don't.

So we look at data edits and then really matching processes. What is appropriate for the type of data that's being looked at, deterministic, probabilistic? Often times it really starts with that positive identification, that data capture point, as well as error tolerance. What is the risk associated with the decision and the usage of that link data? Is it a life or death or safety risk? Is it a legal risk? Is it, quite honestly, someone might have to spend an hour of overtime going in and fixing something? And really, where should those thresholds be set? What is appropriate? In every part of the organization that answer is going to be slightly different and how that's applied really has to be looked into. That really applies to the differences in sizes of the organization as well.

One of the areas, although we focus primarily in financial services, we have gotten involved into the revenue cycle management of healthcare. We've built a relatively large data asset of healthcare information and we've really investigated and applied some of our learnings from linking of records in our financial service products to our healthcare products. A few things that have really come out of that is that even with some of the largest provider networks the data collected in healthcare is not of the same completeness. It's not of the same quality. Our assumption on that is; there's no proof of this; that it's primarily driven because a lot of data collection up front doesn't tie to financial performance and people collect what they need to do; what they have to do today and their goals are not the same as people in financial services.

That being said, there are many ways to link rely on that completeness of the data. All of them rely that it be captured up front as well as it can be. So we really feel that there is a lot of opportunity for establishing at least basic data standards for data capture up front. What gets into the healthcare system, the better is up front. The less issues are propagation of all of the other areas that come along, as well as ensuring that there are processes and systems to support what those basic standards are, the detection, correction of incomplete data and the communication that that culture has to exist.

Further lessons to be applied to get more into the mass techno/geek side of it really are in the areas of probabilistic matching, using demographic data—

Deven McGraw – Center for Democracy & Technology – Director
You've got like four seconds.

Timothy Boomershine – Fair Isaac/FICO (Finance)
—demographic data and behavioral data, as well as, again ensuring that the risk and the usage of the data are aligned with the cost involved in doing that linking.

Paul Egerman – Software Entrepreneur
Great. Thank you, both, very much. Great presentations. We're going to have some questions for you. Actually, I was going to start. Your comments, Tim, were really excellent and I read through your written testimony. One of the things I saw in the document that was interesting is you made the observation that in the financial industry there really is no significant variation in terms of quality of the data you received. I mean the sources are pretty similar and that's quite a bit different from healthcare. I spend a lot of time thinking about why that would be. This is very interesting and ... would like to be where you are.

My question for you or possibly an observation is what occurs to me as very different is on the financial side the consumers see the data. They see it electronically. They see it in monthly bank statements or credit card statements. They're motivated to look at it and they certainly have motivation to call it out if it's wrong. So that, I suspect, is one of the reasons why the financial industry got their data right is they don't have any choice. If they don't get it right the consumer is not going to go to whatever that organization is. Is that a fair statement or do you disagree with that?

Timothy Boomershine – Fair Isaac/FICO (Finance)
I believe that's a fair statement. I think it's actually broader than that. The financial industry in general has a large incentive to get it right up front. I would say it's better quality than what we see in healthcare. It's more consistent quality, good quality. There are a lot of people who would state whether their credit report is good or not. But the financial industry, because they have dollars at risk, they have an incentive to get up front.

The consumer also, once that initial relationship has been established, has that similar motivation that it is their financial health, their information. In the culture we have right now with identity theft and things like that people are much more aware of what is going on with their bank statements, going on in their credit card statements, the Internet; they're able to get on and see what is going on and correct things very quickly and that helps quite a bit. So generally, I think they have it coming both, from the corporate and the origination side of any relationship and then that continues throughout that relationship.

Paul Egerman – Software Entrepreneur
That's very helpful. Also, the first question for you, Laurence, which is you talked about the management of what you do, which is very impressive, the number of people. You have all of these data sources that you've got to match up against. You have some algorithms or secret sauce or something, how you figure out who is who? How do you do that?

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
Well, no. Actually, training is probably the secret sauce I would say. The lowest common denominator in that interaction as far as CBP is concerned is our officer at the border. They're the first person you contact. They're the first person who has an opportunity to either verify that the information you're providing matches not only you, but matches whatever information we might have in the system. They're also the first one to detect information that you might be providing that is inaccurate. It's not that people are trained to determine when people are being disingenuous or anything, but there are questions that you can ask, just general inquiry, how was your trip, things like that, stories that get inconsistent quickly; that leads you to ask other questions. That's where you'll identify when someone isn't giving you accurate information. I mean certainly, you can look at the picture on the document and you can look at the face you're seeing and ideally, it will match and if it doesn't it doesn't.

The other thing I would say is most of us when we're confronted with, frankly, an armed individual we tend to be more inclined to tell the truth. I mean I don't want to under sell that aspect of it. I mean there are advantages to law enforcement that aren't necessarily enjoyed.

Paul Egerman – Software Entrepreneur
I was going to say it is quite a bit different when I think about it that your people are armed, although I did recently have an encounter with a surgeon who had a knife, but it was still a different circumstance.

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
Right. I mean I wouldn't say that no one ever threatens or anything, but the reality is it's a uniformed officer and most people are naturally a little bit more circumspect and they, "Wow, what is the right answer?" So they'll get nervous. They'll say more than they need to say. They'll do all sorts of things.

Paul Egerman – Software Entrepreneur
But I got the sense that you also, as part of your process, inquire to other databases of people –

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
Well, we did.

Paul Egerman – Software Entrepreneur
My question is how do you know if you've got the right person. What is the secret there? Maybe you can't tell me that, but—

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
No. I mean the secret with the other database is it is basically not really a very strong secret. What it is is basically we look at your credential, your travel document, whatever you can provide. Normally what we're pulling from it is your name, a full name, and your birthday. That's the essential match that we're starting with. We're looking to see if there's data in the system that matches that, all right? Now, we all know that that's not adequate given the several trillion people who live on this planet and given that there are spelling variations. I mean you've alluded to the question of variations in spelling and certain alphabets are more phonetic than others. That's not always adequate, so the next step is basically we would have an officer where there is a potential match; we would have the officer inquire of the individual for additional information, other data sources. I mean I won't say it's an elegant process. It's basically a manual process that sort of wears down in the sense of we try to go across a lot of different data points based on what the individual can provide when they're at the border. I mean most reliable, obviously, would be the government issued passport or something along those lines and we would try to go.

Remember, the data when we're confronting you at the border the data we're typically looking not so much to match you to, but finding if there is a match is generally derogatory information. It's not you've just won $1 million; step over here and get your picture taken. It's really going to be there's a want or a warrant; there were previous instances where you crossed and misrepresentations have been made or there are other matches.

Paul Egerman – Software Entrepreneur
You're matching just on name and date of birth? Is that what you're saying?

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
Typically, through text, which is our principal system at the border for screening people as they come across; we're looking at name and date of birth. Yes.

Paul Egerman – Software Entrepreneur
Okay. Do people have questions? We'll start with you, Adam.

Adam Greene – Office of Civil Rights – Senior HIT & Privacy Specialist

Tim, you talked about financial incentives for the financial sector to make sure information is correct. Is there a legal obligation to ensure that information is correct? Most specifically, I'm trying to recall with the Red Flags Rule if there was also other regulations at the same time that set forth obligations on users of credit reports to correct information if they find a mismatch or if they find outdated information.

Timothy Boomershine – Fair Isaac/FICO (Finance)
Right. If a credit granter establishes that they have incorrect information they're obligated to correct that information in any of the public databases. Generally, the way that happens is through the consumer themselves recognizing and alerting both, the credit agency, the credit bureau or the credit granter that they have incorrect information. The dispute process for credit bureaus you hear about quite a bit where the credit bureau will then go back to the credit granter and ask for confirmation. Generally, the onus is on the consumer to prove that the information is incorrect, but there is a responsibility once it's established as incorrect to have to correct it. So that certainly helps in keeping credit information corrected.

As far as information that is stored at the corporate entity themselves, that's generally considered proprietary information for that industry. They consider it trade secret or whatever and they can decide, depending on the level of factual versus opinion type of information that's really up to them to determine.

M
A couple of questions for Tim: I've always sort of thought that the primary linking number for credit reporting and the like was Social number. Is that the case? Is that—?

Timothy Boomershine – Fair Isaac/FICO (Finance)
Social is the primary thing for credit bureau linking. When we get outside of credit bureau data, so when we start looking at demographics data, we start looking at non-traditional credit bureau data, we have to expand outside of the Social and then when we start looking across a bank, for example; a lot of the large banks, as you know, have multiple operating units. They have disparate systems. They have disparate IDs that we then have to go beyond the Social number and come back in.

The other problem with the Social, even when we go to a credit bureau match, there is an issue with the Social similar to the issue if you go, for example, to the European nations where people are creating numbers on the street. Just because you tell me it's your Social doesn't mean it's your Social—

M
Certainly not.

Timothy Boomershine – Fair Isaac/FICO (Finance)
On the other hand, there is a lot of other information I can ask you that's self-reported, for example, on a credit application that I can match against the information in the bureau associated with that Social that will allow me to essentially establish a more positive—

M
So if I were to go and ask for credit though of a bank or to buy a car and they were to ask for my Social number and I would say no I've always been told that that would be grounds for the credit to be denied.

Timothy Boomershine – Fair Isaac/FICO (Finance)
Almost all of the time. Yes.

M
So therefore, the credit granting organizations have developed a de facto standard in the Social number as sort of a way to link individuals to their credit history.

Timothy Boomershine – Fair Isaac/FICO (Finance)
That is the idea, yes.

M

Okay. I understand there are issues with it, but that is sort of one of the things that they rely on as an identifier.

Timothy Boomershine – Fair Isaac/FICO (Finance)

Right. If we look at essentially the primary market of credit that's the other markets of credit, the non-conforming, non-traditional where either someone doesn't have a credit history, doesn't have a Social, whatever the case is and there is still a desire to lend there. Then we tend to get into non-traditional credit bureau information. In the end, a bank is going to want some kind of credit information for most lending or the products.

M

I guess another question to follow up on that: Who regulates? I know the answer to this, but I just sort of want to make sure I understand completely. Who regulates the way that credit bureaus and financial institutions verify credit?

Timothy Boomershine – Fair Isaac/FICO (Finance)

It is regulated under the Fair Credit Reporting Act and it is generally administered by the Fair Trade Commission.

M

Okay. Then another question is, were you here all day today?

Timothy Boomershine – Fair Isaac/FICO (Finance)

For most of the day, yes.

M

Hearing everything that was discussed today and the understanding that we, as you said, I think you said that we don't necessarily have the best quality of data within healthcare. Just random observations, how do we improve this state of affairs? If you were to take banking as a better model because you have better data, what would you say, what are some of the observations you would have as to things that might be used to improve?

Timothy Boomershine – Fair Isaac/FICO (Finance)

The primary areas really are data capture. Ensuring that the fields are filled in—

M

I'm sorry. What?

Timothy Boomershine – Fair Isaac/FICO (Finance)

Ensuring that all of the fields are filled in. I mean I know that's very, very basic, but you would be amazed if you take a record of 40 million visits how many times the name just isn't even there. So ensuring the data is there, ensuring that it is relatively accurate. Some of the things that were talked about, address verification and going a step further and saying not only is it an address, is it a valid address? Is it not the address of the emergency room?

M

Interesting.

Timothy Boomershine – Fair Isaac/FICO (Finance)

So we have essentially the way I would prioritize it is establishing what those standards are, what needs to be filled in, establishing both processes and systems to ensure, to support that and neither of those can make sure that it's 100% and then really going that next step. I believe I heard it mentioned earlier in the VA; they've made a real effort to do this; that you have accountability down to the level of the employee, who's taking that information that they care the data is accurate. They understand why it's important that it's accurate and there's accountability if they decide not to care.

That's the front end of it. Once you get past that if you have as accurate and as complete of data as you can get, as is legally permissible, all of that stuff, you have all of these great algorithms that were talked about with whatever limitations they may have. You have all of these great algorithms that now have a reasonable shot at giving the most likely match that will be available.

M
Great. I guess one question for Mr. Castelli. It sounds like, unlike healthcare—where we're trying to reduce the number of false positives, we're trying to eliminate the false positives as best as possible where we have the wrong person matched. I'm assuming with respect to border information it's okay to have a lot of false positives. In so far as if the terrorist is John Houston and the seven-year-old's name is John Houston and the 40-year-old's name is John Houston I'm assuming you would match them all and say I'm going to check them the next time they come to the border. Is that a fair statement?

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
I would say that it's fair to say that when we're going to make an error we're going to make an error on the side of having more false positives because staffing at the border is such that we have personnel. Our ability to essentially resolve the false positives is present when you're right there. Then having resolved that false positive we have processes, automated processes where tying your identity to your travel document we can ensure the false positive doesn't exist again.

M
Right.

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
So I mean I don't want to suggest that we purposely invite a lot of false positives because, obviously, from a mission perspective the time we spend interacting with a false positive is the time we're not spending looking for someone truly of interest. But yes, I mean I think it is fair to say that if you had to make a choice as to what side of the coin you're going to error on you're going to error on the side of maybe being over inclusive.

That being said, to the extent that we can, the officers are encouraged, for instance, you mentioned yourself versus the seven-year-old boy with the same name. They're encouraged to use the data that is apparent. For instance, this is true, I've had instances where we've had to resolve questions where father is a pilot; the son is a teenager; the son has had some unfortunate interactions with local police involving recreational uses of certain substances. Why are you stopping senior and not junior? Well, it's pretty obvious he's senior and not junior. If you look at the record, you'll see that there's the age discrepancy. So there's a way to resolve that and there's a way to resolve that more quickly. Sometimes you can resolve it just with data. To be honest, I think we prefer to resolve it when the person is present because it's a lot easier to confirm identity when you have the person standing in front of you.

M
Great.

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
I would echo Mr. Boomershine's comments that data collection is really the most important thing. I mean one of the things that we drill our officers on is just making sure that they get it right. That they know we have systems to automatically scan passports for instance. Although, to be honest, in a lot of cases we're still what we refer to as fat fingering data in, using the keyboard to move it into the system. Even then the training is that there are checks to make sure that you're entering it properly because each time you cross the border it does inform the next time you cross. It may inform it in the sense of leave him alone. We don't need to worry about this. Move on. Find someone more interesting. Or it may inform them to check on this issue. So to the extent that you can be accurate it's very important.

Deven McGraw – Center for Democracy & Technology – Director

I would suspect that the accuracy issue matters maybe more to the individual, who is false positive than maybe to the agency.

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
Yes. One of the things we try to stress to the individuals is we recognize that it matters to them. We have processes and we have actually officers who are specifically passenger representatives essentially, who are, at least in the case of the airports, whose role it is to sort of mediate that process, to assist individuals where there's a false positive. But it is a concern to the agency because, as I said, that time spent dealing with a person—I mean no one wants to interrupt travel, you know—

Paul Egerman – Software Entrepreneur
It sounds like a reputational issue too when you pull the seven-year-old kid aside and ask him why the heck—

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
Well, yes, it does make the press. When someone who is under the age of five is identified as someone on a terrorist watch list it raises questions.

Paul Egerman – Software Entrepreneur
Are dirty diapers different than underwear—?

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
Yes. Right.

Paul Egerman – Software Entrepreneur
Joy.

Joy Pritts – ONC – Chief Privacy Officer
I have a question about what the customer does when they find out that the information has been mismatched. I don't think it's a unique circumstance in the financial field or in the health field to end up with records associated with your name. The dispute is not over I really paid that bill on time. The dispute centers on that's not my record, that's not my information. How do you prove a negative? How does the customer prove the negative, that that's not their information?

Timothy Boomershine – Fair Isaac/FICO (Finance)
Generally if it's on a decision around lending or opening an account of some kind, positive identification. So essentially we're looking in that case positive identification is what has to end up happening because almost all financial relationships are voluntary most banks aren't willing to take your word for it they have it wrong. When there's a mismatch usually, similar to the case of a false positive here or a false negative, whatever the case is, I think false positive is actually what we're describing more here. When there's a mismatch, additional data will almost always resolve it, something that differentiates yourself from the mismatched data.

When we get into a mismatch where it's occurring at a different level, either further into a relationship or where you have intentional fraud that's a little bit of a different animal. Because that's not just a data error, but someone is representing themselves as yourself or a customer at that point that's when it kicks in a lot of financial institutions' fraud department. It's very inconvenient for everybody. For you, as a customer, you tend to find yourself in a bad situation in terms of accessing credit, accessing stuff until it's been resolved. The person who is representing themselves as you hopefully will get caught, but most of the time doesn't and so it's not very well thought out yet as to what happens other than some of your larger institutions will go out of their way to try to help you.

Joy Pritts – ONC – Chief Privacy Officer
So I see us as facing that as an issue as medical identity theft is growing because it involves not only financial risk, but also health risk where people are getting healthcare under somebody else's Social Security Number or something of that nature. I'm wondering; I was hoping you might shed some light on

how you resolve those issues when it is absolute fraud, because the stakes are a little bit higher when it's your health than when it's just your money. I mean the money is bad enough, but it's an added complication.

Timothy Boomershine – Fair Isaac/FICO (Finance)
Right and one of the things when we talk about matching and I purposely stayed away from that issue is intentional fraud is something that once it occurs the reaction of the organization that's being defrauded and the person who has the mismatched record, one, there has to be detection to begin with. After the case that detection almost always is going to have to depend on the patient or the person, who is having their information used that's not them, seeing that that's happening, which will usually happen when the bill shows up in the mail, if it shows up and they haven't managed to scoot around that as well. The reaction then at that point really is going to depend on the healthcare provider determining what they're going to do in terms of correcting the records. Similar to the bank is going to have to figure out how much they want to correct the records that are on the financial side of it, how much they really want that person to pay anyway.

What is more ideal is to utilize again different toolsets and these are pretty prevalent throughout the financial services side of it, which detects that fraud up front. So if someone comes in with that Social Security Number and is getting healthcare under a fraudulent Social Security Number or something that's not theirs, just matching that Social Security Number and running something that says that matches the name. That's what you said your name is probably isn't good enough. There needs to be additional matches. I think at that point we're looking more into border patrol here that says what else do we know about this person that this name matches. Does it match the person who is standing in front of me asking for healthcare right now?

What does seem to occur in a lot of the data that we have reviewed in those cases that were marked as fraud is that if that extra step happens the service wouldn't have been provided. It would have been tagged as fraud right up front. But at the clinical end of it, where the service is being provided; generally in the large hospitals, not in the small provider offices; there isn't an incentive or a desire to necessarily stop that as opposed to just keep the workflow moving and move them through. It becomes a business office problem which is well after the fact.

Joy Pritts – ONC – Chief Privacy Officer
Yes.

M
....

Joy Pritts – ONC – Chief Privacy Officer
... with the answer, but—

Paul Egerman – Software Entrepreneur
It's a fascinating issue. I mean I'm thinking about the issue too. It also strikes me that the financial industry and credit cards in particular have all kinds of algorithms they use to detect fraud sort of like after the fact. I don't think healthcare really has ever thought to implement that kind of an approach. So you look at things like the location where services are being provided as being unusual for that individual. There's a constant alert on financial fraud. There isn't that same alert on—

Joy Pritts – ONC – Chief Privacy Officer
I do have a follow-up question, which is you had mentioned address verification as being something that was useful. How useful is it?

Timothy Boomershine – Fair Isaac/FICO (Finance)
Address verification is useful in two functions. One, it ensures you have standardized addresses and ensures it's an address that exists. From the business office perspective one of the things in one of the earlier testimonies was that business office demographics tend to be better than clinical demographics.

That's true, because you can't mail someone a bill if you don't have a good address, so it ensures that it's a good address.

The second part of that is establishing; and this is one of the things that you'll see in financial services; a negative address database basically that says if you've given me this address I don't believe you. The example, I mentioned this earlier, but it is a real example we've observed is in some large, urban hospitals there are thousands and thousands of people who apparently live at the emergency room. I don't know what you do with that. That's the address that you have. That address probably shouldn't be allowed. If we don't have an address there needs to be some kind of standard default home list indication, whatever the case is. So it is important in both of those factors.

Paul Egerman – Software Entrepreneur
It's interesting you talk about the address being an emergency room. It's not that easy if you're in the emergency room and the patient comes in and you're bleeding. You push them through as fast as you can. You've got to keep in mind or at least we all need to keep in mind that lots of time in healthcare the patient doesn't really want to be there. They certainly don't want to be in the emergency room, but even if they're seeing their physician they may not want to be there. They may have a screaming child with them. There may be a line. They may be quite ill. So getting data and all of the data fields filled out neatly and cleanly isn't always that easy under those circumstances. Maybe that's a rationalization, but—

M
... issues as well in an emergency department that prevent us from necessarily taking information down at the time of entry.

Paul Egerman – Software Entrepreneur
Peter.

Peter DeVault – Epic Systems – Project Manager
I was going to respond to your last comments there, Paul. I think that is a difference. It's often led to be a difference that in healthcare if you don't collect the information healthcare can still take place; whereas, if you're applying for credit or opening an account if you don't collect the information it's not going any further, right? The system won't let you or the process won't let you. I have worked with healthcare organizations who don't let that be an impediment or at least to the extent that they can, where the patient is already past the reception. He's being seen by the doctor, but before the patient leaves there will be further information collection, so that really is a business process that needs to be in place. Not every healthcare organization obviously does that, but some organizations do decide that regardless of the clinical outcome the patient isn't leaving until we do collect that information.

Paul Egerman – Software Entrepreneur
Wes.

Wes Rishel – Gartner, Inc. – Vice President & Distinguished Analyst
I think this case really gets back to a discussion we had this morning, which is about data standards. Keep in mind that data standards are more than just formats. So yes, we would like to have a standard, for example, HL-7 allows you to separate the name into three parts. It allows you to separate an address into logical parts. But it doesn't state, as far as I know, within the standard; well, version three does. Version two doesn't state how to say data not present.

In fact, what we hear is that there are some emergency departments that when there is no ability to collect the data they put in, because the system requires there to be data they put in the address of the emergency department. But if we were to say that the standard includes identifying data that's not known and translating that as a not known value then the systems could be smarter if we're willing to put the pressure on all of these hospitals and practices to change their data collection practice to live up to that standard. I think generally one of the things that people complain about in the CDA for it being complex is the fact that it's got half a dozen different ways to represent data not present, but they're like we didn't

ask. We asked; they didn't answer. There are things like that that, in fact, is meaningful information that you could have in lieu of the data.

So standardization, when I heard them calling for standards this morning I think a lot of it was not calling for a new standard format, but calling for the practices to fill out the format properly. We get a lot of cases where they don't collect the middle name or do collect the middle name or deal differently with hyphenated names or deal differently with the former name of a spouse than a middle name and things like that. So just a lot of areas where yes, we have a standard, but we don't have a standard for feeding the standard and that creates these problems that we're describing here.

Paul Egerman – Software Entrepreneur
Excellent comments. Very helpful. I'm sorry. Go ahead, Dixie.

Dixie Baker – Science Applications Intl. Corp. – CTO, Health & Life Sciences
I guess this is more addressed towards Tim than Laurence, but how do you back out errors in a credit report? Like if somebody else's credit information gets into my credit report; this is a common thing that happens in EHRs, right, what is the process for backing that information out or is it backed out or is it just labeled as inaccurate?

Timothy Boomershine – Fair Isaac/FICO (Finance)
The process is fairly straightforward and it's fairly highly regulated. If you identify an error in your credit report, which either, especially if it's data that's not yours, you can either go straight to the credit reporting agencies and they will got to whoever put that information on your report and ask for confirmation. If they get confirmation that it is not you it will be removed, deleted, gone. Some of them are very good. It takes about a day or two. Some of the credit bureaus might take a month or two. From the point you dispute it it will be marked as disputed on your report until that resolution is established. Generally speaking, that may or may not be as successful as a lot of people would like it to be.

The other direction for removing errors in your credit report is to go directly to the organization that put it out there, establish with them that it's on your report from them and you're not that person. As soon as they know and acknowledge it's an error they're legally obligated to have it removed from your report. When it's removed it's gone, deleted.

Dixie Baker – Science Applications Intl. Corp. – CTO, Health & Life Sciences
And there's no note that says anything has ever—

Timothy Boomershine – Fair Isaac/FICO (Finance)
No note. No nothing. The record is wiped out.

Paul Egerman – Software Entrepreneur
John Houston.

John Houston – Univ. Pittsburgh Medical Center – VP, Privacy & Info Security
To follow up on two points, I think that there was discussion about the Red Flags Rule. I think that those types of regulations actually give healthcare a unique opportunity to validate data. Because when an individual comes into the healthcare setting their identity needs to be validated to comply with the Red Flags Rule, which I guess has been delayed again, if I'm not mistaken, but nonetheless, I think there are some opportunities to improve data input through compliance with other regulations. I think that's just one of the comments I wanted to make.

The other one is still that in an emergency room setting especially, you still have ..., which says you must stabilize before you can even really ask much information. So there are going to be times when, yes, the address to the emergency room is what's there or there's no address, simply because they're not medically stable and therefore there is no intake other than to get them into the emergency department itself. So there are reasons behind some of this and I think there are some opportunities as well.

Paul Egerman – Software Entrepreneur
Gayle, go ahead.

Gayle Harrell – Florida – Former State Legislator
One more question on how you handle things on a credit report. For instance, if you eliminate that information completely is there an audit trail or something that tells you that that information has been given to other people from that credit report, to another credit card company or something so that that information can then be deleted from that other source, that other vendor or someone who got that information? Do you have an audit trail system so you can follow that and then make your correction down the pike? Because what happens with health information, it could go down the pike and we need to know where it's been and that it's been corrected.

Timothy Boomershine – Fair Isaac/FICO (Finance)
Right. That is a difference between credit reporting; when it's eliminated from your credit report it will be correct going forward. To get a little off topic, if you pull and look at your credit report you will see everybody who's ever requested it and when they requested it. If you also see something inaccurate it would be up to you if you wanted to contact all of those people and tell them that it's inaccurate, but the credit reporting agencies themselves will—

Gayle Harrell – Florida – Former State Legislator
So the onus is on the client, the customer—there's no responsibility to follow that back?

Timothy Boomershine – Fair Isaac/FICO (Finance)
That is correct. Yes.

Laurence Castelli – Homeland Security – Privacy Officer, Customs & Border Protection
I would note that for the government the onus is on the government to actually maintain a record of accounting when we share your information and when we amend it we inform those who we shared it with that it has to be changed. The Privacy Act requires that, so that's what we would do.

The other thing I was going to mention, Mr. Rishel, I think noted that the procedure was also something that you could look at in terms of how do you standardize data collection. I think that he's quite right in that regard. When I talk about the training our officers receive, part of it is to establish a procedure for how you're going to take information from people and how you're going to fill out forms and how you're going to process them essentially. I readily acknowledge that if someone is bleeding in an emergency room there are certain procedures that need to be addressed first, before we deal with information; however, having a set of practices that are consistent will enable you to improve data quality.

I think one of the differences probably that you see in the healthcare industry that you don't see, obviously, for an agency like Customs and Border Protection is just the fact that we are one agency. My ability to ensure that everyone is following a consistent rule is partially because I can affect the training that they all receive. I can also influence the discipline procedures if they don't comply. I'm sure that that's probably not quite the same thing for you. So I don't know. There are ways that market forces can have an impact on that, obviously. I think that the point about the business office tends to have a better address than the provider office is simply because the market force says you won't get the money if you don't have a good address. So there is that. I don't know if we try to build that in—

M
That's certainly one of the things we noted when we worked through this data asset. I noted in my written that there's a significant difference in the quality across the spectrum of the healthcare data. In particular I go back to the emergency room example; in urban areas we see a lot more of what is most likely you come in, get treated and either don't want to give data, can't get data, whatever the case is and lack of follow-up to get that after the fact. When we get to suburban and rural areas we don't see that same level in the same type of patient, the same type of visits, that same level of inaccurate data; that either it's collected from a relative or it's collected after the fact. But one of the primary drivers or at least correlated factors is that suburban/rural has a much higher financial accountability for the services that were

provided and received than the urban area. That's being driven primarily from the business office. Even for these emergency room visits where in the urban area clearly there is no follow-up to try to collect these bills and the fact that we're coming in on revenue cycle management that part is very obvious as well and that there isn't an financial incentive for most of those particular visits to ever be paid.

Deven McGraw – Center for Democracy & Technology – Director
It does make you wonder if, as we move to different insurance options and we move to getting more people under coverage and people have defined financial obligations that come with that coverage whether some of that—putting care that today isn't in the revenue cycle management might move there down the road, depending on how well this all works, right?

Timothy Boomershine – Fair Isaac/FICO (Finance)
Certainly one of the things that was mentioned earlier is the idea of in financial service having a high amount of fraud detection analytics that go on. Certainly, on the payer side of healthcare some of the gentlemen that were here earlier, there is a lot of fraud detection analytics that exist in that part of healthcare already; on the provider side not very much, if any. What application there is to move that over and start looking at that even at the provider side of is this person really the person they're saying they are, particularly through the link in of healthcare records is an area of opportunity. I'm not really the person to speak to the details of it, but certainly at some point if anyone is interested I can point them in the right direction.

Paul Egerman – Software Entrepreneur
Great. Do we have any other questions from the members of the team here? I don't hear any other questions, so I want to say thank you very much, both of you. Excellent presentations. Extremely helpful. Thank you.

Unless the any member of the team wants to make any other comments I think we're going to go ahead; we're running a little early, but we'll—

Deven McGraw – Center for Democracy & Technology – Director
We have to have public comment and then I wanted to mention some next steps.

Paul Egerman – Software Entrepreneur
Okay.

Judy Sparrow – Office of the National Coordinator – Executive Director
Do you want to do public comment now?

Paul Egerman – Software Entrepreneur
Yes.

Judy Sparrow – Office of the National Coordinator – Executive Director
Okay. Let's take this opportunity to invite the public to make comment. Go ahead, please; your name and your organization?

W
I'm ... I'm a member of the ... Barbara referenced early in her comments. Four quick comments. Number one, as Barbara and I kind of laughed, we think back a few decades ago when we were in health information school, then called medical records school; there was what was called UHDDS, the Uniform Hospital Discharge Data Set. States have that type of minimum data sets also, so we would suggest to the group here there are some federal standards. There are state standards out there for data collection. Now, whether they're met or not, clearly is a different story, but there is history and precedence for monitoring and managing data quality, so you might want to take a look at that. I know several states that have rigorous edit checks for completeness, for thoroughness before you submit to states like New York or California for hospitals or for emergency departments.

Number two; there are health information organizations today that do enforce minimum data quality standards before you're allowed to share data within those HIOs. I think of one I specifically talked to, probably three years ago and their membership, as a part of their governance process enacted that you had to have a certain level of data quality before you joined. Specifically before you shared your data because for the organizations within that HIO who had invested for years or decades in high quality data they wanted to ensure that the data that was going to be shared from other organizations had that same level of investment in data quality so that everybody reaped the benefit of that data quality.

The third comment I would make is around a little hint was made a number of times about the paper-based world. So for Barbara and I, we've been in this profession a long time. In the paper-based world of health information and data sharing, while it was done in a manual form, records are lost. Records are stolen. Records are abused and misused. Privacy is breached. So these are issues that obviously there's a lot of discussion around today, but they aren't new issues. It's something that the profession has faced for a long time.

Then the fourth comment: In some documents that Barbara brought forward and I'd reference about five years ago Dr. Scott from IBM and I did some NCDHS testimony about the completeness of data and what that means to linking of records. So there are some public documents out there that have specifically addressed what completeness means potentially to the matching and linking and sharing of records. Thank you.

Judy Sparrow – Office of the National Coordinator – Executive Director
Thank you for those comments.

Paul Egerman – Software Entrepreneur
Thank you very much. We actually read some of the articles that you wrote. They were circulated and we appreciate your contributions.

Judy Sparrow – Office of the National Coordinator – Executive Director
Are there any comments on the phone? No. All right. Deven, back to you.

Deven McGraw – Center for Democracy & Technology – Director
Terrific. We have a Tiger Team call tomorrow morning from 10:00 to 12:00 and then the Policy Committee meets on Monday. Our goal for the Policy Committee is not to try to present consensus recommendations, because I think we don't quite have enough time, but what we do hope to present to them is sort of a summary. Paul talked about this in the beginning of the day today, a summary of what we heard and getting some feedback from the Policy Committee members that we can then fold in to a recommendations discussion.

So our very helpful folks from MITRE have been, during this whole, entire hearing, helping to sort of take some notes about some common themes. So that we can have a discussion tomorrow on what we will want to put before the Policy Committee, both in terms of here are some common themes that we heard, as well as here are some questions that it might be helpful to get feedback on. Now, keep in mind we only have 15 minutes in the Policy Committee schedule, so there's not a lot of time for dialogue, but it's nice to have an opportunity to tee that up. We won't be able to, I don't think, distribute stuff to folks in advance, but we will have some materials that we can use during the call tomorrow and all of this will be fresh in your mind since we just talked about it today.

Was there anything I missed?

Paul Egerman – Software Entrepreneur
No. I think the only thing left to do is, again, to thank everybody, so I want to, again, thank Judy Sparrow and Joy Fritz from ONC. I agree with Deven. I want to thank the people from MITRE. Of course, thank you to the members of the Tiger Team, who flew from all over the country to be here and to the members of the public, who participated in the call. Thank you very much. Hopefully, you have a chance to join our call tomorrow at 10:00 eastern time.

Public Comment Received During the Meeting

1. Having followed this subject for four years at CMS I am extremely grateful for the excellent collaboration today.   So very much welcomed to get the policy and standards going before folks like me who have to work on the engineering end develop solutions for the Affordable Care Act. Please keep up the pace and effort.

2. CMS would be interested in exploring with ONC the notion Microsoft mentioned of using two certification tokens rather than fuzzy logic.

3. CMS is looking into building Master Data Management style systems.   Panel 2 described the 'Registry Style MDM Pattern' and the centralized style sometimes used in transactional systems.   I suggest CMS and ONC talk to these panel members about the importance of MDM and the 'practices' the one panel member spoke to.   Also, Sara Temlitz has experience at VA (where she works at times with DOD) and would have some insights.   All the while knowing the challenges of Medicare and Medicaid are distinct - there are clearly ideas and practices that can be added to the CMS go forward strategy for related topics.

4. Rather than a hard coded UPI, what about a 'virtual UID' that EMPI networks would use for internal purposes?   This gets to the point one of the speakers raised about stealing them and selling them on the streets.   Ms. Temlitz talked about what VA does. VA uses a virtual id on the back of the VA Card and internally in the database.

5. Do your matching models / studies distinguish between medical records vs. medical claims?   I am with CMS and need to 'integrate' claims, encounters, and assessment records from other Federal Agencies as well as States. Any thoughts?